

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Access DB# 074171

Requester's Full Name: Amy Nelson Examiner #: 73840 Date: 8/22/00  
 Art Unit: 1638 Phone Number 306-218 Serial Number: 08/1984 099  
 Mail Box and Bldg/Room Location: 9F17 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention:

Inventors (please provide full names): McBrat F Al

Earliest Priority Filing Date:

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

PLEASE SEARCH SEQ ID NO: 1, 7, 11, 12, 15

PLEASE RUSH!

FOR OFFICIAL USE ONLY

## STAFF USE ONLY

## Type of Search

## Vendors and cost where applicable

Searcher: D-26-02-02  
 Searcher Phone #: 308-4242  
 Searcher Location: CM112F18  
 Date Searcher Picked Up: 9/1  
 Date Completed: 9/18  
 Searcher Prep & Review Time: 4  
 Clerical Prep Time: 3  
 Online Time: 6  
 NA-Sequence (N) STN  
 AA-Sequence (N) Dialog  
 Structure (N) Questel/Orbit  
 Bibliographic Dr Link  
 Litigation LexisNexis  
 Fulltext Sequence Systems Compuserve  
 Patent Family WWW/Internet  
 Other Other (specify)



Db 18215 ATATAAATATCTATTAAATTTATAATAGTATATAGCTTTTTTAAAAAAAATTAATT 18156  
Qy 2084 ttgacaataatttttgacaataaagttttctaggagataacggaataattctt 2143  
Db 18155 TTTTAAAAAATTTTTAAAAAATGAAAAATAATTAATATATCTTCAATATAAAT 18096  
Qy 2144 ctttttgtaaaataactactgcagaacacaaacgctttgggggcgaataactagc 2203  
Db 18095 TATTTATTAATAATTTTGTGTATTTTAAAAACATGATTTATATATAAATTT 18036  
Qy 2204 ttaagtagtcagtgtaactctgcaactctggtctgaactctcaggtcgagttgctg 2263  
Db 18035 TTTA--TAAAAATATACATTTAAGAATTTTAAAAATTTATATTAATTTATTAAT 17978  
Qy 2264 ctacagtagtaagtcatagaacttaacctgacaaacgacatgcagtcaggctgacat 2323  
Db 17977 AATTTAATTTTCTATATATATATATATATATAAATATCAATATATAAATTTA 17918  
Qy 2324 tacaactttcttttcttccataacttaacttggttgatcaggttcgactataata 2383  
Db 17917 TAAATATATATATATTAATAATTTATATATTTTATAAATTT--AAATTAATATTA 17859  
Qy 2384 ttttaacgattatcaatttcaaatccttatcatctctatataaataaagtcagt 2443  
Db 17858 ATAAATAGGAATATAAATTTTATAAATATATCTACATTTTAAATTTTAAATTT 17799  
Qy 2444 tcaattcgttttgcgaagttcccaaaattttgatttttaaaattatccctaaaa 2503  
Db 17798 TTATTAATTTATAGATATATAATAATATTAATTAATTTATATATATAAATCTA 17739  
Qy 2504 cogaatagttatctttccaaatttaagtttcttttccatccgatttccattctc 2563  
Db 17738 TTATTTATATTTAGTATATAGTTTTTTTTAAAAAAAATTTATTTTTTAAAAAAT 17679  
Qy 2564 cttttataactctctattctatataactacataaattccaaatttttgaatat 2623  
Db 17678 TTTTTTAAAAATGAAATATAAATAAT--TATTTTCATTTAAATTTATTATTAA 17622  
Qy 2624 caatttgctccttaagtcgaactataaatttccatttgaataaactattttcc 2683  
Db 17621 AAATTTTGTGTATTTTAAAAACATGATTTATATATATAA--TTTTT 17569  
Qy 2684 atctaagctcaaaatttaaccaatgcacaataatcagtagtagatcaagcttt 2743  
Db 17568 ATAAAAATATTCATTAAAGAAATTTTAAAAAATTTATATTAATTTTAAAAATTT 17509  
Qy 2744 gaggcttccaaacataaaattacaaaaaaacaaactaaactatttcaatttg 2803  
Db 17508 AATTTTCTATATATATATATATTAATAATTTCAATATATATAAATTTATAAAT 17449  
Qy 2804 aacacaaagcttgccgagatgcgaagctaaagctggtcttttggttctttgt 2863  
Db 17448 ATAAATTAATTAATTAATTTTAAAAAATAAATAAATAAATTTTATTAAT 17389  
Qy 2864 tgcacaagctgggagagaagagggaatgaagattgacacattttttattgtttaa 2923  
Db 17388 -----TAAATTTAGTAATAAATAATTTTAAATTAATTTAGTTAATAATTTAT 17338  
Qy 2924 catataataataaatttaactataaattatactttggtgagtgacgctggggata 2983  
Db 17337 TTATTTATTAATAAATTAATAATGATAGTTTATTTAATTAATTAATTAATAATTT 17278  
Qy 2984 cgtaaagatttttaacattatacttttgcagcgttggtgctgctaccgaagctg 3043  
Db 17277 ATTTATTTAAAAATTTAAAAATTTATTTCAATTTAAATATATATATATATATAT 17218  
Qy 3044 caaagtttgagctgctcgaatgcgaacttttggccataat--ggataaagcaat 3100  
Db 17217 AATTTTATTAATTTTAAATATTTTATTAATAAATTTATTAATAAATAATTT 17158  
Qy 3101 tgttttgctcgaactgctcagcaataagtttaaaatgaatataaagtggtcgt 3160  
Db 17157 TATTAAGTATATTTTAAATAATTTTATTTTAAAAAATTTTATTTTAAAGTTTAA 17098

Qy 3161 cacacacacaaaaaaactaatgtgtgtgtgaattttatatatcggagatgaat 3220  
Db 17097 TATATATAAATTTTATGAATAGGGGGAATAAATTTATTTTCATTTACATATATATA 17038  
Qy 3221 tatattttaaaataaattatgtatttgactcttaatttttgagacattccata 3280  
Db 17037 TATATATATATACAAATTAATTAATTCAGATTTAGTCATTAATAAATAATTTTATTA 16978  
Qy 3281 taatttcgtacataataataataatagta-ataaagtgtaacttaacttaattac 3339  
Db 16977 TACTTATATATTTTAAATGAAATTAATTTATGTATATATATAAATATATGAATG 16918  
Qy 3340 agacataataataatttgactcaatttttttctctatttttttaatttag 3399  
Db 16917 AATTTTATAAAAAATCAATTTAAATTTTATATATAAATAATTTTATTAATTTAT 16858  
Qy 3400 tctattttttccaaataaatttaacttaataaataaatttttcttaattgtgaa 3459  
Db 16857 TAAATTAATTTTATTAATAAATTAATTTATTAATAAATTTATTTAAGTATTAATTTAA 16798  
Qy 3460 caactcagtttactctcaaatatagaatfatatttacttgcgtatttttatta 3519  
Db 16797 TAAATCAATTTTTTTTAAAAAAAATTTTAAAGTTTAAATTAATAAATAATTTAT 16738  
Qy 3520 gttatttaattcgtatttaattatgtgggatacagctcttcccaataatttta 3579  
Db 16737 GAATAGGGGGAATAAATTTTTCATTTTATATATATATATATATATA-ATTAA 16679  
Qy 3580 ctatgatttataaatttttccacatcgtattttactatttaacataatttctac 3639  
Db 16678 TTATTCAGATTTAGTATTAATAAATAATTTTATTTATTAATTTATATTAATTAATGA 16619  
Qy 3640 natttttaggaanttgagcagaacacacttaagagacaaactctatacagaagacat 3699  
Db 16618 AATTAATAATATATATATATATATATAA-----TATAAATGAAT 16574  
Qy 3700 ttgaaaaaagatctattttagtaatttttaagttactcttaacaaacagaagatcaa 3759  
Db 16573 TTTTAAAAATTTTAAATTTTATTAATAAATAAATTTCTATTATTAATTTTAA 16514  
Qy 3760 atcaaatgactaaataagataataacatacaggacatcttacttgtaatttcaact 3819  
Db 16512 ATAAATTTATTTATAAATAATTTATATATAAATAGTTTATTAGTATAATTTAATA-- 16454  
Qy 3820 cccataatttttattatgaanaatacttattatctcgaactaaattgtgtccaaatt 3879  
Db 16455 --AATAATTTTTTTTTTAAAAAATATTTTTT---AAGTTTAAATATATAATAA 16403  
Qy 3880 attactaaataaagaaaacacttaattttataacatttttctataatttgaagat 3939  
Db 16402 TTTATGATAGGGGGAATAAATTTTATTTTCAATTTACATATATATATATATATATA 16343  
Qy 3940 tatattttgtattttacgtaaaattttgacatagattgagacacttcttaacata 3999  
Db 16342 TACAAATTAATTTAGATTTAGCTATTAATAAATAATTTATTATTAATTTATTAAT 16283  
Qy 4000 cccacataagtcagtagtgatgagaattgtgacaaacagctggggccaaatccc 4059  
Db 16282 TTAATGAAATTAATTTATATATATATATAAATAATATGAATGAATTTTATGAA 16223  
Qy 4060 accaaacacatctctctctctctctataaagctgtgc--tacacatagacaacatcca 4117  
Db 16222 AAATCAATTTTAAATTTTATTAATATAAATAATTTTATTAATTTTAAATAATTTT 16163  
Qy 4118 cacacaataacagctctttcttctctatttgataacacagtcgctcgaactgtcacc 4177  
Db 16162 ATTTATAAATAATTTATTAATAAATAAGTTTATTAAATTAATTTTAAATAATTTT 16103  
Qy 4178 cettctctctcttcccaacttactcagaagctgcactagt----ggccgttagcc 4232  
Db 16102 TTTTAAAAAATAATTTTAAAGTTTAAATATACAAATAATTTATGAATAGGGGAA 16043

```

RESULT 15
DMU37541/c
LOCUS      DMU37541    19517 bp    DNA    circular    INV      04-APR-2000
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION  U37541
VERSION    U37541.1  GI:1166529

```

KEYWORDS	Drosophila melanogaster.
SOURCE	Drosophilinae Drosophila melanogaster
ORGANISM	Bukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 12511 to 12682)
AUTHORS	Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.W. and Wolstenholme,D.R.
TITLE	Drosophila mitochondrial DNA: a novel gene order
JOURNAL	Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE	83090428
REFERENCE	2 (bases 5269 to 5695)
AUTHORS	Clary,D.O., Wachtelthier,J.A. and Wolstenholme,D.R.
TITLE	Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes
JOURNAL	Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE	83220794
REFERENCE	3 (bases 404 to 5272)
AUTHORS	de Bruijn,M.H.
TITLE	Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code
JOURNAL	Nature 304 (5923), 234-241 (1983)
MEDLINE	83245048
REFERENCE	4 (bases 804 to 1778)
AUTHORS	Satta,Y., Ishihara,H. and Chigusa,S.I.
TITLE	Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species
JOURNAL	Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE	88174373
REFERENCE	5 (bases 5268 to 13619)
AUTHORS	Garesse,R.
TITLE	Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations
JOURNAL	Genetics 118 (4), 649-663 (1988)
MEDLINE	88212147
REFERENCE	6 (bases 441 to 2957)
AUTHORS	Satta,Y. and Takekita,N.
TITLE	Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE	91088557
REFERENCE	7 (bases 14215 to 14512)
AUTHORS	Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.
TITLE	Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods
JOURNAL	Science 258 (5086), 1345-1348 (1992)
MEDLINE	93088057
REFERENCE	8 (bases 14917 to 19517)
AUTHORS	Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE	Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA
JOURNAL	Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE	94285822
REFERENCE	9 (bases 1 to 408; 13319 to 19517)
AUTHORS	Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE	Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons
JOURNAL	Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE	96423163
REFERENCE	10 (bases 1 to 19517)
AUTHORS	Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE	Direct Submission
JOURNAL	Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES	location/Qualifiers
source	1..19517 /organism="Drosophila melanogaster" /organelle="mitochondrion" /db_ref="taxon:722?" /note="Derived from new and previously submitted



```

FEATURES             Location/Qualifiers
     source            1..67970
                        /organism="Plasmodium falciparum"
                        /strain="3D7"
                        /db_xref="taxon:5833"
                        /chromosome="1"
gene                 complement(1748..3276)
                        /gene="MAL1P3.01"
CDS                  complement(join(1748..2598,2748..2848,2990..3276))
                        /gene="MAL1P3.01"
                        /note="MAL1P3.01, conserved hypothetical protein, len: 412
aa, similarity: UPF0006 family eg to
YBLOS5C/YBLOS12/YBLOS11, YBFS_YEAST (418 aa), fasta
scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
overlap)"
                        /codon_start=1
                        /product="conserved hypothetical protein, UPF0006 family"
                        /protein_id="CAB63556.1"
                        /db_xref="GI:6594244"
                        /translation="MKLVFHYIKYINVLFTISIIPLASLNKSIYNDLRYISTVYKVKY
LQIKRSNLKKNENIRMEDNESSPIDIGSNLTDKMPDGVTSKSGKHGENDQLWNLAKR
NNVNDKIITLCTCAEDLSKLCICTYDPGKFLYLSAGVPTNCTEFDIKNHEKEKE
IIAAKEYEZFPIKFNQEVNSKMGNGKICDGEKDDNNLWELLKNNLDTPGFPT
NEXDKREYENLANKIKYIENRIVCGIEIGLDPRLYFPCSTYQIKTFIPQLKLVQMF
LPMFLHNNKSCETTFPIVDIVYLFKFNKGWGHSPTSDIEDIVHYIVQYKNLYIGVG
CSLKLEENINAVKIPLLMLLLELTDAPMGVKKTHSYEYIKDTEKRAATNNLKKIKN
IIRKDDPIPERNEPNYNA"
misc_feature          complement(2599..2610)
                        /gene="MAL1P3.01"
misc_feature          /note="potential splice acceptor sequence"
                        complement(2742..2747)
                        /gene="MAL1P3.01"
misc_feature          /note="potential splice donor sequence, atg/gtaaaa"
                        complement(2849..2861)
                        /gene="MAL1P3.01"
misc_feature          /note="potential splice acceptor sequence"
                        complement(2984..2989)
                        /gene="MAL1P3.01"
misc_feature          /note="potential splice donor sequence, aaa/gtaaaa"
                        5005..5496
                        /gene="MAL1P3.02"
gene                  5005..5496
                        /gene="MAL1P3.02"
CDS                  /note="MAL1P3.02, hypothetical protein, len: 163 aa,
contains possible signal sequence"
                        /codon_start=1
                        /product="hypothetical protein, MAL1P3.02"
                        /protein_id="CAB63557.1"

```

```

/db_xref="GI:6594245"
/translation="MKLLNRSFVFLVCPILILFFPLNSVLGNNNRNINRHETENAAK
AMKRLISGEINSLDGLDKRIKLDNRKHDSKTQWDSSTFISLSEEEKSQYDLPFR
KQEIENENKTIEDRQEPYLIWNDEIENIATRFVLNPFDELYITQSKQSLDIIGS
LN20"
misc_feature
locus_0.10389
/note="possible cen1, region of very high [A+T] content"
gene
14884. 20352
/gene="WALIP3.03"
CDS
14884. 20352
/gene="WALIP3.03"
/note="WALIP3.03, putative ABC transporter, len: 1822 aa"
/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63559.1"
/db_xref="GI:6594245"
/translation="MTTKEKNCVGLSNKGNKKKSQIQNLPSLFDWIRPLINDLIK
QUGELPNCANVDVPTASKLEENLADVEDSDSPTSEKSSNEVHLHSCNSDSS
KVIVHTNHTLASLITKFKRILILISFYLETILVLAGKQITDTHMLRQCPYPI
TSLFQDLKQVSGVSLVVMIFMLFPLBALHAFLEPITLMKWSVMTILKILNCN
NLIKWNQAFPIVTHKFSSTQETIDEDISDFSLIGNASSSSGSIKNNNKIDNKAFF
DIT11HFISIKNNKSGSLNEMRGLPWNITNIMFSDSPSYTFVYVLCILNPNVFPY
MSTFVTHIKKKSQSVIAISLIALISAMLEPFLSPFKSLITLTKRADAIDNKHVIL
KEFKLIRKNSDWSFAKTIINFMKEMTKCRILRPLSGVYFSSISDISEVVEVPIF
LTKRDNKKEEIDETISIMPLVYTKLISLNVAFLPNVNNWGVNIVKRLAWTIDIE
LFTYDITAKIMHTIRKQDILNIVDVTFLQENITSHDDQTSNHLKELANVIAKLTLS
KEPIFFTHKNNKHNKI11NQLISGLNVDVDDTNKIKCFQEKHSKSNYTNVSSHIE
KKESTNTHHNSNSTNSFEKAKNNNEY11KLEKSCFGLSTDKDCKNDKILKINPIL
LRNLSA111IGNVSGSKAPFSLSDGPNMTNGHLY1IMFPAKMP1LVPQNSWLMG
NISMILFGNHYMLPYKTYTILQSELNADTISIEBGDMT1NDHSLSGN1PCCN
RATLYEYTHMKILTDYKKLQPNEDLKDILNKN1NKNSSVSNKSLVNTNIPVLA
TLMCLFEDNLYL1LDDPISLSDPSISKRIFSNLCEKNSIDFDMCSP1ISMAKS
YDGLD1LEDLQWQYEVNIPEDQTKATRYGN1SEYMEKNNNKLTKESHGVS1NLTN
DEKTRK1KLFDEVELNKGNSNMKTYKATFVAPGVSVSEF1DS1NKET1KKMKKAK
KEEMKNNKNNNNNNNNKDDIN1NNNDNWRKYNIDMLGNPDDSDPVS5SGNE
T1D1T1SSNDSKE1KVLPT1KDTHEKSSSSSPVSSVSS1NMP5NF1K1EDN55
FKGSL1ST1LTFQWQGVPLT5V5PML1S1PDEIK1EFTN1N1N1N1N1N1N1N1N1
LQTKQVRYLETVP1L1S1LTVSG1CF5M1TYGNT1SAK1VNN1L1S1NAP1LPT1
YNNNG1N1K1R1ID1SADF1GPKRIYKAPF1PFC1LSS1L1N1N1N1N1N1N1N1
L1LTVYFVKRFSFGKCEAQLYLSCTP1C1N1Y1SNALSG1N1N1KNY1N1N1N1
HY1N1F1N1S1P1R1K1IN1WASLY1K1P1ILL1T111IMPL1AS11K1K1K1N1V1
L5T1G1C1FSFAR1G1Y1FL1C1D1T1HEKEM1CQVLEAP1EAK1SKNKA1NNK1N1
N1Y1T1Q1T1K1EN1SD1SA1VEYK1N1SS1INS1QD1G1S1G1N1K1N1N1N1
KAT1P1N1N1L11DEEPS1K1NN1MY1AK1N1Q1K1V1G1SG1S1G1S11L1GL1N11S1G1
K1T1VEGR1D1Y1TK1R1K1G1S1G1L1G1A1Q1S1F1V1N1N1N1N1N1N1N1N1N1
G1N1K1N1K1L1Y1K1H1G1K1D1M1S1Y1K1K1I1Q1S1K1V1N1Q1SD1N1T1L1N1D1C1R1L1S1V1L1N1
K1R1K1I1L1D1E1P1F1N1L1N1W1D1G1SL1G1K1A1K1S1P11TR1N1H1P1N1V1L11SH1N1T1
LSCDD1ITV1LKE1G1ET1Y1R1C1SD1Y1K1T1Q1E1SL1H1L1E1D1"
rRNA
28966. 31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large
subunit rRNA genes and spacer regions"
gene
28966. 31533
/gene="rRNA"
complement(31966..32775)
/gene="WALIP3.04"
CDS
complement(join(31966..32476,32675..32775))
/gene="WALIP3.04"
/note="WALIP3.04, conserved hypothetical membrane protein,
len: 203 aa, similarity: P. falciparum chromosome 2,
PF0810W, O96126 predicted integral membrane protein (255
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
in 191 aa overlap)"
/codon_start=1
/product="conserved hypothetical membrane protein,
WALIP3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKSYTFPIWITLFLFLLFLITWDTFSKTFKNFNKIDIN
TRFKI1AAGEASQKTPWEEDPCLLIEEELIPENADSDPLPHEHTENDKINELNSIN
TFRK1ETIKRNGYIKETDNNKWRDPWGYWKNATYIKYVHLKINITIKDOLZVPI
HDEKETITTMK1WQDIEYFLFPLQVEMRLILLEIPYKKE"
misc_feature
complement(32477..32486)

```

\* 76337 76316: gap of unknown length  
\* 76317 77913: contig of 1597 bp in length  
\* 77914 77993: gap of unknown length  
\* 77994 80808: contig of 2815 bp in length  
\* 80809 80888: gap of unknown length  
\* 80889 82776: contig of 1888 bp in length  
\* 82777 82856: gap of unknown length  
\* 82857 85682: contig of 2826 bp in length  
\* 85683 85762: gap of unknown length  
\* 85763 89308: contig of 3546 bp in length  
\* 89309 89388: gap of unknown length  
\* 89389 93939: contig of 4611 bp in length  
\* 94000 94079: gap of unknown length  
\* 94080 109466: contig of 15387 bp in length  
\* 109467 109546: gap of unknown length  
\* 109547 110183: contig of 637 bp in length  
\* 110184 110263: gap of unknown length  
\* 110264 110842: contig of 579 bp in length  
\* 110843 110922: gap of unknown length  
\* 110923 111636: contig of 714 bp in length  
\* 111637 111716: gap of unknown length  
\* 111717 112030: contig of 314 bp in length  
\* 112031 112110: gap of unknown length  
\* 112111 112674: contig of 564 bp in length  
\* 112675 112754: gap of unknown length  
\* 112755 113301: contig of 547 bp in length  
\* 113302 113381: gap of unknown length  
\* 113382 113979: contig of 598 bp in length  
\* 113980 114059: gap of unknown length  
\* 114060 114698: contig of 639 bp in length  
\* 114699 114778: gap of unknown length  
\* 114779 115187: contig of 409 bp in length  
\* 115188 115267: gap of unknown length  
\* 115268 115938: contig of 671 bp in length  
\* 115939 116018: gap of unknown length  
\* 116019 116539: contig of 521 bp in length  
\* 116540 116619: gap of unknown length

Query Match 3.4%; Score 189.2; DB 55; Length 161891;  
Best Local Similarity 40.1%; Pred. No. 1.5e-13;  
Matches 1471; Conservative 0; Mismatches 2156; Indels 42; Gaps 14;

Qy 1800 ttaataattattataaattatgacttttgacgtctgactaattttcgaatt 1859

Db 139978 TTAATAATAATAAATAAATAATTTTGTATATATGGAATGTATATATATATATATA 139919

Qy 1860 ttatttttggtttgggtttgttgatatttttttagataatttttttaattatgctaa 1919

Db 139918 TATATATATATATATATATATTTATATATATATATATATATATATATATATATAT 139859

Qy 1920 tttttctgtatttgaaaggatgttcyaatttttttttcaaaatgaacgttttaagat 1979

Db 139858 TTGATATATAAATATGATATATAATATATATATATATATATATATATATATATAT 139799

Qy 1980 ttttactactgcaattcagaataagtgattgttttttgaagaagattaaatagtag 2039

Db 139798 ATTATATATATATATATATTTATGAATATATATATTTTAAATATATATATATATAA 139739

Qy 2040 tattaagatttttttggtt-atttggtggaagtatgattgttttttgacataattatt 2098

Db 139738 TAAATAAGAGATAAATATGTGTATATGTATATATATATATATATATATATATATAT 139679

Qy 2099 tgacataatttaagttttctagggaataacggaatatctctctcttttttgaaatt 2158

Db 139678 TGGAT 139619

Qy 2159 actaetgacgaagacaacacagctttggggagcaataactagcttaagtagtcagt 2218

Db 139618 NNN 139559

Qy 2219 taactctcaaatctgctgaactcttaacgttgctgtgtgctgaactagtagtc 2278

Db 139558 NNNATATATATTTATATATATATATCA--TTTATATATATATATATATATATATAT 139501

Qy 2279 tatagaacttacctgcaacacacagcatgcgtcggtcgatctacaacttttccitt 2338

Db 139500 TAAATAAATAATTAATTAATTTATATATTTTATATATTTATATATATATATATATAT 139441

Qy 2339 tcttcaattcaaatatgctgctgaacttcgcgatctataatattttatcagttat 2398

Db 139440 TTAATAATATATATATATATATATATATATTTTATATATATATATATATATATATAT 139381

Qy 2399 caatttcaacttcaacttatatactattataaataagtcagtcagtcagtttctg 2458

Db 139380 ATATGATATATATATATATATATATATATATTTTATATATATATATATATATATATAT 139321

Qy 2459 aaagtctccaa-aaatttgatttttataaatttttccctaaacgaatagttata 2517

Db 139320 ATACTTTTATATATATATATATATTTTATATATATATATATATATATATATATATAT 139261

Qy 2518 tcttcaaattttgatttgcatttttcaacttcagttcaatttcaacttttataactct 2577

Db 139260 TATTATGTATATATTTTTTATATTTTAAATTTAAATATATTTTTTATATATATATAT 139201

Qy 2578 tattataataataataaattcaaatatttgaatttgaatttcaactttttagctct 2637

Db 139200 TATTATATATATATATTTTATATATATATATATATATATATATATATATATATATAT 139141

Qy 2638 agttcaaacactataaatttttcaattttagaattatcaatttttccatcatagctcaa 2697

Db 139140 TTTTATATATATATATATAT-----ATTATATATATATATATATATATATATAT 139091

Qy 2698 tttaacaaatgcacaaatttcatgattagtagtagcagacttttgacttcaaaaca 2757

Db 139090 TAAATATATATATATATGAAATATTATAGTTTATATATATATATATATATATATAT 139031

Qy 2758 taanaattcaaaaaaaacaaacttaaaacttttcaatttgaacaaacagcttg 2817

Db 139030 ATATATATATATATATATATATATATATATATATATATTTTATATATATATATATAT 138971

Qy 2818 gcgaatgctgaagagcttaaaatgctcttttttcttttttcttttgcgaacgttgag 2877

Db 138972 ----ATATTTAAATTTATATATATATATATATATATTTTGTATCTATCTACTGCAAT 138917

Qy 2878 agaaagggaagtaggaattgacatttttttattatttttcaacataaataaatt 2937

Db 138916 TAT 138857

Qy 2938 aatttaataataattatttgggtgaatgacagtgaggagatcgttaagattttta 2997

Db 138856 TATAATAT--AAATAAGATTATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 138798

Qy 2998 acattataacttttgcagacgttgctgtctcaacgaagatcgaagtttgagctg 3057

Db 138797 NNN 138738

Qy 3058 ccttcaatgcagcatttttttccataatggataaagcaattttttagtctcaactgt 3117

Db 138737 ATATATATATATATCTAT 138678

Qy 3118 cacaagaataagtttaaaatgaataaataaagtggtcgtgctgcacacacaa----- 3172

Db 138677 AT 138618

Qy 3173 aaaaactaaatggttgggttgatttttatatacaggaagtaatttttatattttaaaa 3232

Db 138617 TAT 138558

Qy 3233 taanaattatgtattttagatttttaatttttggagacttcaactataattttgtaac 3292

Db 138557 CAT 138498

Qy 3293 ataatatttaaatattgataataaagtgtaattcaacttaacttaacagataattata 3352

Db 138497 TATAATGTAT 138438

Qy 3230 aaaaataaaatgttatttaggtattctaatattttggagatctcaactataattgt 3269  
Db 100070 AACTTAAATTAATTAATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 100129  
Qy 3290 aacataaataaaataagataataaaagtgtaactaaactaaatacaagaataata 3349  
Db 100130 TAAATATTGTTTATTATATATATATATATTATTAAATTATTATTATTATTATATA 100189  
Qy 3350 taaattttggaacaaactaaattttttctctatttttaattaaatttagctattttt 3409  
Db 100190 TTAAT-----ATTATTAAATTTATAAATAAATTTAAATATATAAATTAATTATTAT 100244  
Qy 3410 caaaataaaattaaactcaaaataaaataattttcttaagtgtgaacaactgaatg 3469  
Db 100245 TATATATTAAATTAATTAATATATATTATTAATTAAATTAAATATATATTATTATTAATTA 100304  
Qy 3470 tatacttcaaaattaaagtatattatacttaccgtgatgtttttatttagtatataat 3529  
Db 100305 TTTTATTATTATTAATTAAATTAATATATATATTAATTAAATTAAATATATATATTAA 100364  
Qy 3530 tctgataataaattatggtgggaacaactgcttccactaaaattttcaactatgatta 3589  
Db 100365 TTTAATATATATATTAATTAAATTGAATGTATATATTAAATTATTATATATATTATT 100424  
Qy 3590 taaattattttcaactctgtatacttacttaattaaacaattatcatcaattttatgg 3649  
Db 100425 TAAATTATTATATATTAAATTATATATATTATTATTATTTTTATTAACATATAAATTATTA 100484  
Qy 3650 aaattggagacaaagaacatttaagaagacaaaattctatacaagaacaaatttgaaaaa 3709  
Db 100485 ATTTAATTAATTTATTATTATTATTAATCAATAATATATTAATTATATATATTATTG 100544  
Qy 3710 atgtactttttaggttaatttttaagtaactcttaaccaacaacaataattcaataaaga 3769  
Db 100545 TTTATTATTATTAATTAATTCATCTTATTAAATTATTATATATTATTATTAATTAATCT 100604  
Qy 3770 cttaaataagataatacaatacagcgaaactcttaactgtaacttcaactccataatt 3829  
Db 100605 TTAATTATTAATTAATTAATGTTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 100664  
Qy 3830 ttaattgaaaataaactctataactctgaacaaattggtgcacaataattatactaaa 3889  
Db 100665 TAAATTTAATTAATTATATATATTAATTAAATTAAATTATATATATTAATTAAATTAT 100724  
Qy 3890 taagaagaaaacactcaatttttaacacttttttccatattatgaagatataatttgt 3949  
Db 100725 TATATATAAATTATTAATTATTAATTAAATTAAATTAATTAATTAATTAATTAATTAATTA 100784  
Qy 3950 atattacagcaaaaaatttt 3969  
Db 100785 TATTAAATTAATTAATTAATTT 100804

```

RESULT 13
AC008206.c
LOCUS      AC008206 161891 bp    DNA             BTG              08-MAR-2000
DEFINITION Drosophila melanogaster chromosome 3 clone BACR03J15 (D765) RPCI-98
            O3.1.15 map 96B-96B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 133 unordered pieces.
ACCESSION  AC008206
VERSION    AC008206.9 GI:7208834
KEYWORDS   BTG; HTGS_PHASE1
SOURCE     fruit fly.
            ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 161891)
AUTHORS   Celnikier,S.E., Agbayani,A., Arcania,T.T., Baxter,E., Blazer,R.G.,
            Belinoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Jhumstis,S.R., Karra,K.,

```

D5	75037	AAATAAAGCTCTATTAATAATTTTTGATAGATTTCTAGATAATATGAATAATTC	74978
Qy	1862	atttttggttttgggttttgttgaaatttttttagataatttttaataattctgcgaatt	1921
Db	74977	GTAAGAAAAATATATATAAAAAATGATATTTAAATATTATATTAACATA-CANATATAT	74920
Qy	1922	ttctctactgcgaattgcgaattgttcgaatttttttccaaattgaaagcttttagaatt	1981
Db	74919	ATTATTTATTATTAATATTTATTATTAAAAATATAAATATTATTATTCGAATTTT	74860
Qy	1982	ttactactgcgaattgcgaattgttcgaatttttttgaagaattcaataagtagta	2041
Db	74859	AAATTAATTAATTAATTAATTTGGTAT-ATTATTAATTTATTAATAAATATTTTAA	74801
Qy	2042	ttacgatttttagttttgttgcgtggaagtatgtagtttttgaactaatttttga	2101
Db	74800	ATATATCACACAAATGATCTTAATTAATATAAAATATTATTATTTATTAAGAAA	74741
Qy	2102	caataaagttttctcaggaaataaacgggaaatactctctctctttttgaaattact	2161
Db	74740	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	74681
Qy	2162	aatcgacgaacaaacacgctttgtggggacaaataactagcttagtagctgtagta	2221
Db	74680	TATTTATTAAATTAATTTATTTATTATTAAATTTAAATAAAATAAATAATATAT	74621
Qy	2222	ctctcaaaattcgtctaatctcttagctgagtttgcgtgcacagtagtagctctat	2281
Db	74620	TATATTATATAATATATTTAATTAATTAATTTATTTATTTAATTAATTAATTAAT	74561
Qy	2282	agaactactcctgcacaaacgacatgcagctcaggtcgaactcaaaattctctcttc	2341
Db	74560	TAAATAAATTAATTAATTAATTTATTTATTAATTAATTTATTTAAATTAATTTAT	74501
Qy	2342	ttcaataacatctggttgatcgaagtcgcagctataataattatcacgattataca	2401
Db	74500	TTTATTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	74441
Qy	2402	tttcaataacctatctatccctatacaaaataagtagctcaactcagtttgcga	2461
Db	74440	AAATATATTCANATTAATTAATTTATTAATTAATTTAAATAAATAATTAATTAAT	74381
Qy	2462	gtccccaacaaatttgagatttttaaaattatccctaaaacggaataattatctct	2521
Db	74380	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	74321
Qy	2522	tsaaatttaagttctcttttccaaatcgaattcaaatctatccctttataactcttat	2581
Db	74320	ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	74261
Qy	2582	atctataataacataaatttcaaatcaatttgaatttcaattcttagtccatagtt	2641
Db	74260	TATTCATATTAATTAATTAATTAATTTATTTTATTAAAAAATAATTAATTAATTAAT	74201
Qy	2642	caaaactataaattttcaacttcagaaataatcaattcttccaaatcgaagctcaaat	2701
Db	74200	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	74141
Qy	2702	accaaatgacacaaattctcgtctagtagaagaagcttttgctctccaacataaa	2761
Db	74140	ATATTTATTTAAATTAATTAATTAATTTATTTTATTAATTAATTAATTAATTAAT	74081
Qy	2762	aattacaacaaacaaacaaactaaactcttataactttcaatttgaaacaaagcttgccg	2821
Db	74080	TATTTAAATTAATTAATTAATTAATTAATTTATTTATTAATTAATTAATTAATTAAT	74021
Qy	2822	aatgcgaagagcttcaaaatgcctcttttctgtctcttttgcgttcaaacggtgagaga	2881
Db	74020	ATAATTAATTTAATGATTATTTATTAATTAATTAATTTATTTATTTATTTATTTAA	73961
Qy	2882	gagggaatggaagttagacatacttttttatattgttttaacct---ataataata	2937
Db	73960	TAAATTAATTAATTTAATTAATTAATTTATTTTAAATTAATTAATTAATTAATTAAT	73901

Query Match 3.4%; Score 189.6; DB 41; Length 104992;  
Best Local Similarity 44.9%; Pred. No. 1.5e-13;  
Matches 979; Conservative 0; Mismatches 1179; Indels 22; Gaps 6;

Qy 1802 aaataattatttaattaaaatttatggacttttggactgtctgactaattttcagaatttt 1861

JOURNAL Science 258 (5086), 1345-1348 (1992)  
 MEDLINE 93088057  
 REFERENCE 8 (bases 14917 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L., Parquhar,A.L. and Kaguni,L.S.  
 TITLE Sequence, organization, and evolution of the A+T region of  
 Drosophila melanogaster mitochondrial DNA  
 JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)  
 MEDLINE 94285822  
 REFERENCE 9 (bases 1 to 408: 13319 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Drosophila melanogaster mitochondrial DNA: completion of the  
 nucleotide sequence and evolutionary comparisons  
 JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)  
 MEDLINE 94423163  
 REFERENCE 10 (bases 1 to 19517)  
 AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,  
 Michigan State University, East Lansing, MI 48824-1319, USA

FEATURES  
 source Location/Qualifiers  
 1. .19517  
 /organism="Drosophila melanogaster"  
 /organelle="mitochondrion"  
 /db\_xref="taxon:7227"  
 /note="derived from new and previously submitted  
 sequences; sequence is a composite containing segments  
 obtained from different Drosophila melanogaster strains"

tRNA  
 1. .65  
 /gene="mt:ND6"  
 /product="tRNA-Ile"  
 /db\_xref="FlyBase:FBgn0013685"

gene  
 1. .19517  
 /gene="mt:ND6"  
 /note="mitochondrial NADH-ubiquinone oxidoreductase chain  
 5"

tRNA  
 /allele=""  
 /db\_xref="FlyBase:FBgn0013685"  
 complement(97..165)  
 /product="tRNA-Gln"

tRNA  
 171..239  
 /gene="mt:ND6"  
 /product="tRNA-Phe"  
 /db\_xref="FlyBase:FBgn0013685"

CDS  
 240..1265  
 /gene="mt:ND6"  
 /codon\_start=1  
 /db\_xref="FlyBase:FBgn0013685"  
 /transl\_table=5  
 /product="NADH dehydrogenase subunit 2"  
 /protein\_id="AAC47811.1"  
 /db\_xref="GI:11656330"  
 /translation="MFNNSKILPITIMIGITLITVTSNWLGAAGLNLINLLSPFL  
 LSDNNNLNTEASLKTFLTVLSTVILFSSILMLKNNKNNNEINESTPMISGALL  
 LKSGAAPPFWFNNMGEGLWNNALMLTWOKIAPLMLISLYNLIKLLISVLIVII  
 GAIGGGLNLSLRKLAFSSINHLGNLSSLMISESINLILFFFSVLSVLFWNFI  
 KLFLHNLQFSNFWNSKILKFTLEFNFLSLGLPLPFLGFLNVLQQLNLQYMLT  
 DMNLSLILFLFYLRICYSAFNNYFENNMKNNNNNNTNNNMIMTFFSLRGLELI  
 SLFYF"

tRNA  
 1264..1329  
 /gene="mt:ND6"  
 /product="tRNA-Trp"  
 /db\_xref="FlyBase:FBgn0013685"

tRNA  
 complement(1322..1383)  
 /product="tRNA-Cys"

tRNA  
 complement(1403..1468)  
 /product="tRNA-Tyr"

CDS  
 join(1470..1472,1474..3009)  
 /codon\_start=1  
 /exception="mechanism underlying reading frame shift after  
 first codon uncertain"  
 /transl\_table=5  
 /product="cytochrome c oxidase subunit I"  
 /protein\_id="AAC47812.2"

/db\_xref="GI:7412849"  
 /translation="MSRQWLESTNKKDITLTFIFGAGAGWGTSLILRAELGHPG  
 ALIGDQDITNVITVTAHAFITFMNMPIMIGFGNVLPLAGAPMAFFRNNNGFW  
 LHPALSLLVYSSVSNVNGAGTGWVTFPLSAGIARGGASDLAITSHLAGISSILGA  
 VNFITVTVMNRSTGSLDRMFLFWSVVTALLLLSLPLAGAITMLTDRNLATSF  
 FDPAGGGDGLYQHLFWFHPFVEVITILLPGFGMISHLISQSGKATFGSLGMITAM  
 LAIGLGLFVVAHMRFTVQMDPITRAITFSAIMIIVAPGKIIFSLATGLSTGLSTS  
 PAILNALGFFVFLTVGGLGVGLANSSVDLILITTTVAHFVLSMGAVFALMGAF  
 IHWYFLPTGLTNKKLNSHFIMFIVGNLTFPPGHFLGAGHPRI SDTPCATTTWN  
 IVSTIGSTLGLLFFFFIINESLVSQRQVITPLQNLNIEWMTQPPAHSSTSELP  
 LLIN"

tRNA  
 3012..3077  
 /gene="mt:ND6"  
 /product="tRNA-Leu"  
 /db\_xref="FlyBase:FBgn0013685"

CDS  
 3083..3767  
 /note="TAA stop codon is completed by the addition of 3' A  
 residues to the mRNA"  
 /codon\_start=1  
 /transl\_except=(pos:3767,aa:TERM)  
 /transl\_table=5  
 /product="cytochrome c oxidase subunit II"  
 /protein\_id="AAC47813.1"  
 /db\_xref="GI:1165532"  
 /translation="MSTWANLGQDSAPMQLQIFFDHALLILVMITVGVGLMFW  
 LFNNTVYRFLLAGGLEIMWITLPAILLFLALPSLRLLILDELMEFSTYKLSILG  
 QWNSYESTDFWNIFEDSTMTPLNMLTDGFLRDDNVRVFLPNQIIRILVTAADVI  
 HSWTVPALGVKDGTPGRLNQNFNFIRPLFGYGCSEICGANSFMPVIESVPVNY  
 FIKWISSNNNS"

tRNA  
 3768..3838  
 /gene="mt:ND6"  
 /product="tRNA-Lys"

tRNA  
 /db\_xref="FlyBase:FBgn0013685"  
 3840..3906  
 /gene="mt:ND6"  
 /product="tRNA-Asp"  
 /db\_xref="FlyBase:FBgn0013685"

CDS  
 3907..4068  
 /gene="mt:ND6"  
 /codon\_start=1  
 /db\_xref="FlyBase:FBgn0013685"  
 /transl\_table=5  
 /product="ATPase 8"  
 /protein\_id="AAC47814.1"  
 /db\_xref="GI:1165533"  
 /translation="MPQAPISLWLLFIIFSTIFLPCISNYTSPMKNPSNELKMI  
 NLNSNNWAK"

CDS  
 4062..4736  
 /gene="mt:ND6"  
 /codon\_start=1  
 /db\_xref="FlyBase:FBgn0013685"  
 /transl\_table=5  
 /product="ATPase 6"  
 /protein\_id="AAC47815.1"  
 /db\_xref="GI:1165534"  
 /translation="MNTNLSFYDPLAIFNFWLNLSTGLGLMISYIWNLMPSRYNI  
 MNNLSILLKPEKTKLPGSGNGSTFIISLSLLENFMGLFPIFTFSGLATL  
 LSLALPNLWCLMYGNINHTQMPFAPVQGTQPALMPFVCIETISNIRPQTAVR  
 LTNAMIGHLILLTGLTSSSMYMLTFLMLAQIALVLASVAMIQSYFVATLSTL  
 YSEVEN"

CDS  
 4736..5524  
 /gene="mt:ND6"  
 /codon\_start=1  
 /db\_xref="FlyBase:FBgn0013685"  
 /transl\_table=5  
 /product="cytochrome c oxidase subunit III"  
 /protein\_id="AAC47816.1"  
 /db\_xref="GI:1165535"  
 /translation="MSTNNSFFHNDVSPWFLGAGCANMTVSGWVNFHQYDLSLF  
 VLGNITLILVYQWNRDREGTQGLYHTAVTIGLRNGLFLILSEVFSVFWAF  
 FHSLSLPAELGASGWPWPGIISFVFPFGLPALLATLACSGVTWNAHSLNNNSGT  
 TQGLFFVLLGIYTLQAYEYIAPPTIADISYGSFTPATGFGHVLIGTITFLVL  
 CLLRLNHNLSKNHFGFAAAYTHFVDVWVFLITTYIYWGQ"

Query Match 3.44; Score 189.6; D9 33; Length 4601;  
Best Local Similarity 44.34; Pred. No. 3.7e-13;  
Matches 1063; Conservative 0; Mismatches 1309; Indels 28; Gaps 6;

Qy 1538 ttttcgaataaatttcgcatgtggtttttcgaataattgcatacaatcatcgca 1597  
Db 2203 TTTTAAAAAAAATATTATTAAATATATCTAATAAATATTTTTATATAAATTA 2262

Qy 1598 tctcgaattttgcgaattgaacgtataaaattctctatgatctcgtatctgtatt 1657  
Db 2263 TTTTAAATAAATAATTATTAATAAATAATTAATAAATAATATATATATATATATA 2322

Qy 1658 acattatagt-tgtttatgcttgagtgtaagtcacaacttgcagattcagtcacccact 1716  
Db 2323 TATTAATAAGAAATTAATTTTAAATTTTAAATAATAAATTTAATATATATAATTA 2382

Qy 1717 tatattacatttcaggcaactcgcagacttagattagctggccttcaggagcttggt 1776  
Db 2383 AATAAATCTATTCTATATATATATTTAATTAATATAAATAAATTAATATCTAATAAT 2442

Qy 1777 tggttttctccatccatttttatataaatttatataaaatttatgaccttttga 1836  
Db 2443 TAATAAATTTATTATTCTCAATATTTAATATATAATAAAAAATTTTTTTTTTTT 2502

Qy 1837 ctgctcg-actaatttccgaattttattttgggtttgggtttgtgaatttttga 1894  
Db 2503 TTTTTTTTAAATATTAATTATATATATATTTAAATATTAATATATATGAATAT 2562

Qy 1895 taattattttaaaattcttcgataaattttctgttatttgaaagagtgctgaattgt 1954  
Db 2563 TTAATATATATATATATATATAGAAAAAATAATTTATTAATAAATTTAATATAAATTT 2622

Qy 1955 ttccaataatgaacgcttgaagaatttttactcgcgaactcagataaagtgaattgt 2014  
Db 2623 TAAAAATTTCTTAAAGTATTTATTTTAAAAAAATTTATATATAAATCTGTTT 2682

Qy 2015 tttttgaagaatttaataa-ctgatatagatttttagtttgatttggtggaagt 2072  
Db 2683 TTTAAAAATAAACAATAATTTTAAATAAATAATTTAATAGAAATATATTTATT 2742

Qy 2073 aagtatgtttttgaaacaaattattgcacataaataagttttctagggaataacgga 2132  
Db 2743 ATTTTTCATTTTAAAAAAAATTTTAAAAAAAATATTTTTTTTTAAAAAAAAC 2802

Qy 2133 aatatctctctcttttttgaataattactatgcagacacacacagctttgggggaca 2192  
Db 2803 ATATCTAATATAAATTAATGATATTTATATATATAAATATTAATATATATTAT 2862

Qy 2193 aataactagatttaagtagtaagtgcaactcgaactcgaactcgtctataactctagctg 2252  
Db 2863 ATATCTAATAAATTTAAATAAAAAATTTTAAATTTAAAAAGTAGATATATTTATAAA 2922

Qy 2253 agtttgctgctcagtagtaagctcatagaaaactacctgcagaaacagcatgaagtc 2312  
Db 2923 ATTTAATTTCCATATTTTATTATATTAATTAATTTATATAAATAATATAAATTA 2982

Qy 2313 agggctgaactatcaaacctttctcttttctccaataacatacgttgctgaactgt 2372  
Db 2983 ATTAATATATATATTTATAAATTTATATATTTATGAATATTTATATATATATATA 3042

Qy 2373 atctataaattatttatcagatttatcaatttccattaccttatactctctattataaa 2432  
Db 3043 TATATGAAAAAATAAATATTATAAATAATTTAAATAAATTTTAAAAAATTTCTTAAA 3102

Qy 2433 tataagtcagtcgaactcagtttgcgaagctcccaaaaaatttgaatttttataaatt 2492  
Db 3103 TGTATATTTTTTAAAAAATTTATATATAAATAAATCTGTTTTTAAAAAATAACAA 3162

Qy 2493 attccctcaaacggaatagtataattctccaatttaagtttcttttttccatccagatt 2552  
Db 3163 AAAATTTTATAAATAAATTTTATATGAATATAATTTATTATTATTTCATTTTTTT 3222

Qy 2553 tcaatttcaactgtttataactctctattatcatataataatcgaataatcgaatttaatt 2612

```

/gene="MAL3P5.7"
/note="predicted using hexExon: MAL3P5.7 (PFC0605c),
Hypothetical protein, len: 248 aa"
/codon_start=1
/protein_id="CAB41709.1"
/db_xref="GI:4725991"
/translation="MGGGGAILNLIPQKKNVWRDQAQKHYHDEHKIKKEKDEKKR
KDEPESITSLAKMKWKNEDSUNNYNFTDENGKAKTITITCNDMHNLFDIEKELT
AQKQKEEFLIKGHYITITDANFNQGNISITDKNNQAQIISDFKNKCELDQWFLP
NNKNETQKNDNPNFHKQDNLSEHRKTHINSDLSLCTVMNITTHDKAKKQKMY
HIKKIKLQPKSKDEKKRKRQGEKKKKP"
complement(29992..33537)
/gene="MAL3P5.8"
complement(29992..33537)
/gene="MAL3P5.8"
/note="predicted using hexExon: MAL3P5.8 (PFC0610c),
Hypothetical protein, len: 1182 aa"
/codon_start=1
/protein_id="CAB38971.1"
/db_xref="GI:4493935"
/db_xref="SPTRMBL:Q97260"
/translation="MAHKVKKKKETAEQETPVVAKETHAKEENNSNIATVENNIS
KMGQETAKSKNDQETAKSKNDQETAKSKNDQETAKSKNDQENVALNSSEKQASKEE
DWELQKQIFGEHDINSNHNENRNSFTSLSSFGFEEKNSVGGFSGTHTDHSME
DSNNVETENAIINDVLRNSRSTSYSKQKNEILSYTCVTCGEVDVNLNSDHLFAHKL

```

- ```
*      1      67262: contig of 67262 bp in length
*      67263      67462: gap of unknown length
*      67463      82485: contig of 15023 bp in length
*      82486      82685: gap of unknown length
*      82686     130281: contig of 47596 bp in length
```

Query Match 3.6%; Score 197; DB 60; Length 130281;  
Best Local Similarity 45.7%; Pred. No. 2.1e-14;  
Matches 1118; Conservative 0; Mismatches 1290; Indels 37; Gaps 11;

Qy\_2270\_tagtaagtctatagaaacttacctgacaaaacgacatgacgtcagggtcgaatctacaac 2329

Qy 3338 acaagcataatattaaattttgaatcaattaatctttatcttattatcttaattatctt 3397



LOCUS 140338 1283 bp DNA PAT 13-MAY-1997  
 DEFINITION Sequence 17 from patent US 5620882.  
 ACCESSION 140338  
 VERSION 140338.1 GI:2082630  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 1283)  
 AUTHORS John.M.  
 TITLE Genetically engineering cotton plants for altered fiber  
 JOURNAL Patent: US 5620882-A 17 15-APR-1997;  
 FEATURES Location/Qualifiers  
 source 1. 1283  
 /organism="unknown"  
 BASE COUNT 509 a 233 c 251 g 290 t  
 ORIGIN

Query Match 5.08; Score 273.4; DB 5; Length 1283;  
 Best Local Similarity 84.24; Pred. No. 2.1e-22;  
 Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

Qy 4124 aatacagctcttctttttttttattgataaacctgctcatgactctgcaaccccttc 4183  
 Db 13 ACTAAAAATCTTCTGCTTCTATTTTGTAAACCATGGCTCATACATTTTGTACATCTTTC 72  
 Qy 4184 ttctttttccaaacttttactaagtgctctcaagtgacagcgttgacacactgttttg 4243  
 Db 73 TTCTTTTCCACATTTTACTCATCTCTCTCACTAATAATGGTAGTCACACCGCTCTG 132  
 Qy 4244 gcagcgctgcagctttatctgagacacagcaacccctcatgagcttcccaactgctt 4303  
 Db 133 TCAGCGGCTGCGATTTATTCGACACACAAACCATCTCATGAGCTGCCCAATTGGCT 192  
 Qy 4304 tcaaatcagaagaagcagagagcttgatgaacgaagcagaatacaaacagcgaag 4363  
 Db 193 TCAAAATACGAAGAAGCAGGAAGGCTCT-----GAATCAAGAACAACAAA 237  
 Qy 4364 tatccagagagctactcaaaacttgagacgctgaagtcaaaagaggaagaaacaaaa 4423  
 Db 238 TATCCAGAAAGTACCCAAACATGAGAGGCTCAAAATGCACAGAGGAGAAACAAA 297  
 Qy 4424 cctgcgaacagcagtgagagctacacagctcacagcactcaaaagagcaaaagagatc 4483  
 Db 298 CCTGCAGACATCATGAAGAGTACACAGAGTACGCGGAATCGAAGGACGACGAAGATAC 357  
 Qy 4484 gagaagaagaactctgcgggcccga 4510  
 Db 358 GATAAGGAAAAACCGGATTCGCCAAA 384

RESULT 6  
 AC005504  
 LOCUS AC005504 104992 bp DNA HTG 01-APR-1999  
 DEFINITION Plasmodium falciparum chromosome 12, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 3 unordered pieces.  
 ACCESSION AC005504  
 VERSION AC005504.3 GI:4558584  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 104992)  
 AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurd,I.O.B., Conway,A.B. and Davis,R.W.  
 TITLE Plasmodium falciparum 3D7 chromosome 12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 104992)  
 AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-NOV-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA

94304, USA  
 COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 58642: contig of 58642 bp in length  
 \* 58643 58642: gap of unknown length  
 \* 58843 91011: contig of 32169 bp in length  
 \* 91012 91211: gap of unknown length  
 \* 91212 104992: contig of 13791 bp in length.  
 FEATURES Location/Qualifiers  
 source 1. 104992  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /chromosome="12"  
 BASE COUNT 44286 a 9326 c 9564 g 4141 t 405 others  
 ORIGIN

Query Match 3.68; Score 197; DB 41; Length 104992;  
 Best Local Similarity 45.78; Pred. No. 2.2e-14;  
 Matches 1118; Conservative 0; Mismatches 1290; Indels 37; Gaps 11;

Qy 1556 atgtgtgtttctgaaatatgcatcaacataatcatgcttctcaatttggctcaat 1615  
 Db 72352 ATTTTATTATTTCATATAAAAAAGGATAAGACATATAAATAATAATAAAAAACA 72411  
 Qy 1616 tgaagctataaactctctatgatactcgtctgtttatcaactatgtgtttatg 1675  
 Db 72412 TTCCAAATATATACCCCCAATATATATATATATATATATATATATATATATATAA 72471  
 Qy 1676 ctgttgatgaagcaacttgagattcatagctcaccoccatatttatactttccgagca 1735  
 Db 72472 TATATATATACATATATATATATATATATATATATATATATATATATATATATATAT 72531  
 Qy 1736 atctgcgacttaggttggtgtggttcgaggttggttggtgtgtttcttccactcat 1795  
 Db 72532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 72591  
 Qy 1796 ttattataaataattatataaataattatgagcttttgagctgtgactaatttcag 1855  
 Db 72592 AAAATTAATATATATATATATATATATATATATATATATATATATATATATATAT 72651  
 Qy 1856 aattttattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1915  
 Db 72652 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 72711  
 Qy 1916 ataattttctgttatttgaagaggtgtgcaatttttttccaaattgaaagctttaa 1975  
 Db 72712 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 72771  
 Qy 1976 gaatttttactctgcaaatctgagaagtgcaattgttttttgaagagtaataaag 2035  
 Db 72772 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 72831  
 Qy 2036 tagtatctagactttt---agttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2092  
 Db 72832 TTAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 72891  
 Qy 2093 atattttgcaataataagttttctagaggaataacggaataatcttcttcttttgg 2152  
 Db 72892 ATAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 72949  
 Qy 2153 aaattactactgagagacaacaacagcttttgggagcaataactagcttagtag 2212  
 Db 72950 TAAATTAATATATATATATATATATATATATATATATATATATATATATATATAT 73009  
 Qy 2213 tcaagttaactctcaaaacttggtcataactctagc---tgagttctgtgtctacg 2269

U49822 *Saccharomyces*  
Z92859 *Caenorhabditis*  
X95275 *P. falciparum*  
AC007708 *Homo sapiens*  
AC008132 *Homo sapiens*  
AL031745 *Plasmodium*  
AC004617 *Homo sapiens*  
AB011369 *Plasmodium*  
U434560 *Plasmodium*  
AC006280 *Plasmodium*  
AL031747 *Plasmodium*  
AC008132 *Homo sapiens*  
X95276 *P. falciparum*  
AB011368 *Plasmodium*  
U49822 *Saccharomyces*  
U434558 *Plasmodium*  
U4034397 *Human DNA*  
AC004617 *Homo sapiens*  
AC004470 *Homo sapiens*  
AC004153 *Plasmodium*  
AC005083 *Homo sapiens*

## ALIGNMENTS

```

RESULT 1
GB034401
LOCUS      GB034401    1699 bp    DNA                PLN           01-JAN-1996
DEFINITION Gossypium barbadense FbLate-2 gene, complete cds.
ACCESSION  U34401
VERSION     U34401.1  GI:1143223
KEYWORDS
SOURCE      sea-island cotton.
ORGANISM    Gossypium barbadense
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosid II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 1699)
AUTHORS     Rinehart, J., Petersen, M. and John, M.E.
TITLE       Tissue-specific and Developmental Regulation of Cotton mRNA,
            FbLate-2: Promoter Studies in Transgenic Plants
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1699)
AUTHORS     John, M.E.
TITLE       Direct Submission
JOURNAL     Submitted (21-AGO-1995) Maliyalkal E. John, Fiber Technology,
            Agracetus, 8520 University Green, Middleton, WI 53562, USA
FEATURES
source      1..1699
            /organism="Gossypium barbadense"
            /strain="Sea Island"
            /db_xref="taxon:3634"
            /clone="FbL2-82A"
mRNA        369..1585
            /gene="FbLate-2"
            /gene="FbLate-2"
gene        369..1585
            /gene="FbLate-2"
            /codon_start=1
            /protein_id="AA084861.1"
            /db_xref="GI:1143224"
            /translation="MIGSYTVSYTAARLFETQTTSSELPLASKYIKQDESSYEKPK
            NKPTDEDEPKRHKPELHKKEKQAPKQHEETHSHSHAKSHHEEYQAKPEFPALEK
            HKHEHVEYTPILTKYKNDQGEKHKHECHSHSHESRSHHEEYERKPEFPALEKPE
            EKHEHVEYPIPEPKYKQDGEKHKHECHSHSHESRSHHEEYERKPEFPFPALEKPE
            HEVYTPILPEYKQDGEKHKHECHSHSHESRSHHEEYERKPEFPFPALEKPEKHEK
            VEYTPILPEYKQDGEKHKHEFPNQKHEEKKPKKATYSENPFSMPTQSGSTPQ
polya_signal 1448..1454
            /gene="FbLate-2"
BASE COUNT  661 a      328 c      328 g      382 t

```

```

RESULT      2
118362
LOCUS       118362      1283 bp      DNA              PAT      07-OCT-1996
DEFINITION  Sequence 17 from patent US 5495070.
ACCESSION   118362
VERSION     118362.1  GI:1598717
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
             Unclassified.
REFERENCE   1 (bases 1 to 1283)
AUTHORS     John,M.
TITLE       Genetically engineering cotton plants for altered fiber
JOURNAL     Patent: US 5495070-A 17 27-FEB-1996;
FEATURES
            Location/Qualifiers
             1..1283
              /organism="unknown"
BASE COUNT  508 a      233 c      251 g      290 t      .

```

```

RESULT      2
118362
LOCUS       118362      1283 bp      DNA              PAT      07-OCT-1996
DEFINITION  Sequence 17 from patent US 5495070.
ACCESSION   118362
VERSION     118362.1  GI:1598717
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
             Unclassified.
REFERENCE   1 (bases 1 to 1283)
AUTHORS     John,M.
TITLE       Genetically engineering cotton plants for altered fiber
JOURNAL     Patent: US 5495070-A 17 27-FEB-1996;
FEATURES
            Location/Qualifiers
             1..1283
              /organism="unknown"
BASE COUNT  508 a      233 c      251 g      290 t      .

```

Search completed: September 2, 2000, 22:58:41  
Job time: 19337 sec

Db 299 TATGGGACCTGACAGGACAGAGATTACATAGGTGAGGCTTTAGTATACAGAG 358  
 Qy 241 ctgagtgttttttggcctttttcttataagcaagcaggttatgaaacatctaca 300  
 Db 359 CTGATGTGTTTTGTGTGCTATCTCTCATCAGCAAGCCGATGTAGAGAACTCTCA 418  
 Qy 301 aaaaagtgtccagagcagcattatgtctaatgtaccctgtgtctgttggaa 360  
 Db 419 AAAAGTGGATACCTGAGCTGAGCATTATGCTCCAAATGTGCTTATGCTGGGGGAA 478  
 Qy 361 ccaactagattgtcgagatgacacagcagcttctcattgatccactggagcaaccca 420  
 Db 479 CAAACTAGATNTGCGAGATGACAGCAATNTCTGATTGATCATCGGGATCGCAGGAA 538  
 Qy 421 tatacaactctcagga 437  
 Db 539 TAACACTGCTCAGGCA 555

RESULT 12  
 A1900170  
 LOCUS A1900170 658 bp mRNA EST 06-DEC-1999  
 DEFINITION sc01g12.y1 Gm-c1012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1012-959 5' similar to SW:RAC1\_PEA Q35638 RAC-LIKE GTP BINDING PROTEIN RH01. ; mRNA sequence.  
 ACCESSION A1900170  
 VERSION A1900170.1 GI:5606072  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Bacteria; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 658)  
 AUTHORS Shoemaker,R., Reim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Oct 30, 1998 this sequence version replaced gi:3812130.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 401.

FEATURES  
 source  
 location/Qualifiers  
 1..658  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1012-959"  
 /clone\_lib="Gm-c1012"  
 /tissue\_type="Apical shoot tips, 9-10 day old etiolated seedlings"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II XR; Site.1: EcoRI; Site.2: XhoI; this cDNA library was constructed from mRNA isolated from the apical shoots of 9 to 10 day old etiolated seedlings. The shoot tips including any emerged leaves were harvested for mRNA isolation. The cDNA library was

prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

BASE COUNT 176 a 125 c 171 g 184 t 2 others  
 ORIGIN

Query Match 33.6%; Score 305.6; DB 45; Length 658;  
 Best Local Similarity 73.9%; Pred. No. 7.7e-64;  
 Matches 386; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 5 aacaactgacagcagcattatcaagctgtgacgtgtgatgagcgtgtgggaa 64  
 Db 136 AAGATGACGCTCTTAGTGTTCATCAAGTGCCTGCTGGGGATGCTGTGGGCAA 195  
 Qy 65 aactgtgtatgcttattctatccagcaactatttccacagcattgttccacagt 124  
 Db 196 AACCTGCTGTCTTATTCTCTACACGACCAACACTTCCACACCGATTATGTGCGCAGTG 255  
 Qy 125 attgataacttattgtccaatgtgtgtgtgtgagcagcagtgaaacctgtgt 184  
 Db 256 TTGTGCAATTCTGAGTCAAGTGTGTGTGCAATGAGGACGATTGTAATCTGGTTGTG 315  
 Qy 185 ggacactgcggcggaagaagattataagctgaagcagcagtgattagagagcgtga 244  
 Db 316 GGATACTGCTGGACAGAGATTACACAGATTAAAGACTTTGAGTTACCGTGGCGCA 375  
 Qy 245 tgtgttttttggcctttttcttataagaagcagcattgaaacatctaca 304  
 Db 376 TGTTTTCTATGATGCTTCTCTCTCATAGCAGGCCAGGTATGAAATGTCTTAAAA 435P  
 Qy 305 gtggatccagagcattgaacattatgtcctaattgtacacgtgtgtgtgtgaa 364  
 Db 436 GTGGATCCAGGCTGAAGCATTATGACCTGTGTCCTCCCAATTATTCTGTGGCAGCA 495  
 Qy 365 actgattgtcgagatgacacagcagcttctcattgatccactggagcaaccca 424  
 Db 496 GCTTGACCTGTGGGATGATTAGCAGTTCTGATGACCACTTCTGGTGCCCTACTATTAC 555  
 Qy 425 aacatctcagggagagaactaagaagatgagtaggagcagcttacttatagatcag 484  
 Db 556 CACAGCTCANGAGAGAGCTTAGAAGCTGATTATGACCACTTACATTGAATGACG 615  
 Qy 485 ctccaaaccccaagaagattgtgagcgtgttttgcattgtcgtc 526  
 Db 616 TTCAAAAACACGAGAGACGCTGGATGAGCTCTTGTATGCTGACG 657

RESULT 13  
 A1901141  
 LOCUS A1901141 549 bp mRNA EST 06-DEC-1999  
 DEFINITION sc01b12.y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1013-1272 5' similar to SW:RACD\_GSH1 Q41253 RAC-LIKE GTP BINDING PROTEIN RAC13. ; mRNA sequence.  
 ACCESSION A1901141  
 VERSION A1901141.1 GI:5607043  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Bacteria; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 549)  
 AUTHORS Shoemaker,R., Reim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,

ACCESSION AI900160  
 VERSION AI900160.1 GI:5606062  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
 REFERENCE 1 (bases 1 to 688)  
 AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpeljing,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Jul 30, 1997 this sequence version replaced gi:2285374.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 431.

FEATURES  
 source  
 Location/Qualifiers  
 1..688  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl012-935"  
 /clone\_lib="Gm-cl012"  
 /tissue\_type="Apical shoot tips, 9-10 day old etiolated seedlings"  
 /lab\_host="XL10-Gold"  
 /note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2: XhoI; This cDNA library was constructed from mRNA isolated from the apical shoots of 9 to 10 day old etiolated seedlings. The shoot tips including any emerged leaves were harvested for mRNA isolation. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpeljing."

BASE COUNT 191 a 132 c 179 g 184 t 2 others  
 ORIGIN

Query Match 33.8%; Score 307.2; DB 45; Length 688;  
 Best Local Similarity 74.1%; Pred. No. 3.2e-64;  
 Matches 387; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 5 aacaatgcagcactgcagattatacaagtgtcctacgctgtgctgagctgtggtggaa 64

Db 166 AAGATGAGGCCCTCTCTAGCTTCACTCAAGTGGCTCATCTGTGGGAGTGGCTGTGGGCAN 225

Qy 65 aactgtatgtctcaattctatacagcgaattactttcccaacgattatgttccacagt 124

Db 226 AACCTGCTGCTTCTTCTTCCACACGACACATTTCCACACGATATGTGCCGACTGT 285

Qy 125 atttgataacttttagtgcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 184

Db 286 TTTTGCAATTTCAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 345

Qy 185 ggcactctggcgccagagattataagctgaagccactgtgtatagaggagctga 244

Db 346 GGATCTGCTGGACAGAGATTACACACAGATTAGACTTTGGATTACGTTGGTGGCGA 405

Qy 245 tgtgtttttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 304

Db 406 TGTTTTCATATGTGCTTCTCTCTCATAGCAAGCGACTTATGAAATGCTCTTAAAAA 465

Qy 305 gtggatccacagactgaacattatgtcctaattgtaccagtgtgtgtgtgtgtgtgtgt 364

Db 466 GTGGATCCAGAGTGAAGCATATGACACGTGTGTGTGTGTGTGTGTGTGTGTGTGT 525

Qy 365 actgattgttgagtagacacagcacttctcattgtaccctgtgtgtgtgtgtgtgtgt 424

Db 526 GCTTGACCTCTGGGATGATAGACGTTCTGATGACACCATCTGTGTGTGTGTGTGTGT 585

Qy 425 aacatctcaggagagagaactaaagaagatgatagagcagcttacttatagaatcag 484

Db 586 CACAGCTCANGAGAGAGACTTAGAAGCTGATTATACCACTGTACATGAATGACAG 645

Qy 485 ctccaaaccccaagaatgtgaagcgtgttttgcagtgtcgt 526

Db 646 TTCAAAACACACAGGAGACGTGAAGGACGCTTTGATGACAG 687

RESULT 10  
 AW705028  
 LOCUS AW705028 592 bp mRNA EST 18-APR-2000  
 DEFINITION sk41f03.y1 Gm-cl019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl019-5142 5' similar to SW-RAC5\_ARATH Q38937 RAC-LIKE GTP BINDING PROTEIN ARAC5. [1] , mRNA sequence.  
 ACCESSION AW705028  
 VERSION AW705028.1 GI:7589250  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 592)  
 AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpeljing,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Jun 22, 1998 this sequence version replaced gi:3246649.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 430.

FEATURES  
 source  
 Location/Qualifiers  
 1..592  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl019-5142"

|    |     |                                                              |     |
|----|-----|--------------------------------------------------------------|-----|
| Qy | 6   | acatatgacactcgaagtattatcaagtgtgtaacggtcgtgtatgagcgtctggggaaa | 65  |
| Db | 794 | AAATCAGTGNNNINGAGTTCATTAACTGGCTCNNNTGNCWGCAGCGTGCTGTCCNNHAA  | 735 |
| Qy | 66  | acttttgatgctcaattcotaatacagcaactctcccaaggatgatgttccacagcta   | 125 |
| Db | 734 | NNNNGNCNNTGATTNNNACACCGACCAACTGTTTCACGCGACTATGTGCCANNNTT     | 675 |
| Qy | 126 | ttgtataactttatgtgccattgtgtgtgtagtcgacgacacagtgaactctggcctatg | 185 |

```

RESULT 4
AW040005
LOCUS      AW040005      732 bp      mRNA      EST      18-OCT-1999
DEFINITION EST282496 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
           clone cLET19J17, mRNA sequence.
ACCESSION  AW040005

```

Qy 365 actagatttgcgagatgacaagcagttcctcattgatcaccctggagcaacaccaatattc 424

```
117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query |        |     | ID | Description |                     |
|------------|-------|-------|--------|-----|----|-------------|---------------------|
|            |       | Match | Length | DB  |    |             |                     |
|            | 1     | 379.2 | 41.7   | 662 | 79 | AW690945    | AW690945 MF034H11S  |
|            | 2     | 339.8 | 37.1   | 595 | 79 | AD1937960   | AD1937960 sc06b1.y  |
|            | 3     | 337.6 | 37.1   | 732 | 63 | AW039993    | AW039993 EST282484  |
|            | 4     | 337.6 | 37.1   | 732 | 63 | AW040005    | AW040005 EST282484  |
|            | 5     | 336.8 | 37.0   | 585 | 71 | AW394676    | AW394676 sh3a40.2   |
|            | 6     | 324.4 | 35.6   | 669 | 74 | AW559248    | AW559248 EST3306084 |
| c          | 7     | 318.6 | 35.0   | 796 | 71 | AW349629    | AW349629 NC0100558  |
|            | 8     | 308.4 | 33.9   | 649 | 44 | AT759963    | AT759963 sb6b61.f   |
|            | 9     | 307.2 | 33.8   | 688 | 45 | AT900160    | AT900160 sc01f12.y  |
|            | 10    | 306.8 | 33.7   | 592 | 80 | AW705028    | AW705028 sbk1f03.y  |
|            | 11    | 306.6 | 33.7   | 555 | 45 | AT901151    | AT901151 sc21c12.y  |
|            | 12    | 305.6 | 33.6   | 658 | 45 | AT900170    | AT900170 sc01g12.y  |
|            | 13    | 304.2 | 33.4   | 549 | 45 | AT901141    | AT901141 sc21b12.y  |
|            | 14    | 303.4 | 33.3   | 713 | 44 | AT759954    | AT759954 sb6b61.f   |
|            | 15    | 299   | 32.9   | 605 | 79 | AW688369    | AW688369 MF060620S  |
|            | 16    | 295.4 | 32.5   | 506 | 46 | AT965741    | AT965741 sc75409.y  |
|            | 17    | 291   | 32.0   | 533 | 74 | AW573665    | AW573665 BST31362S  |
|            | 18    | 290.2 | 31.9   | 549 | 36 | AT162543    | AT162543 AT0199200  |
|            | 19    | 282   | 31.0   | 486 | 46 | AT1941239   | AT1941239 sb6c610.f |
|            | 20    | 278.6 | 30.6   | 805 | 74 | AW573660    | AW573660 EST316251  |
|            | 21    | 277.2 | 30.5   | 469 | 47 | AD029919    | AD029919 AD029919   |
|            | 22    | 267.2 | 29.4   | 517 | 80 | AW705209    | AW705209 sbk43a11.y |
|            | 23    | 266.4 | 29.3   | 622 | 43 | AT727570    | AT727570 BMJL6H1842 |
|            | 24    | 264.8 | 29.1   | 410 | 74 | AW559842    | AW559842 EST314890  |
|            | 25    | 262   | 28.8   | 680 | 79 | AW690086    | AW690086 MF021805S  |
|            | 26    | 261.2 | 28.7   | 435 | 44 | AT1812534   | AT1812534 12DB      |
|            | 27    | 257.2 | 28.3   | 437 | 69 | AW020293    | AW020293 sf31c10.y  |
|            | 28    | 255.4 | 28.1   | 638 | 43 | AT731040    | AT731040 BMJL6H1845 |
|            | 29    | 252.8 | 27.7   | 463 | 48 | AD086932    | AD086932 AD086932   |
|            | 30    | 249.8 | 27.5   | 401 | 40 | AT495724    | AT495724 sb15406.y  |
|            | 31    | 249.8 | 27.5   | 568 | 44 | AT757563    | AT757563 EST25663   |
|            | 32    | 247.2 | 27.2   | 696 | 80 | AW694335    | AW694335 MW0750026  |
|            | 33    | 244   | 26.8   | 698 | 54 | AW105094    | AW105094 cat0002P   |
|            | 34    | 237.8 | 26.1   | 556 | 69 | AW129991    | AW129991 EST302474  |
|            | 35    | 236.6 | 26.0   | 551 | 79 | AW621657    | AW621657 EST312455  |
|            | 36    | 231.8 | 25.5   | 560 | 80 | AW738459    | AW738459 EST3139826 |
|            | 37    | 229   | 25.2   | 688 | 79 | AW685566    | AW685566 MF031896   |
|            | 38    | 227   | 24.9   | 617 | 49 | AW021840    | AW021840 EST303563  |
|            | 39    | 225   | 24.7   | 524 | 43 | AT730323    | AT730323 BMJL6H1842 |
|            | 40    | 224.6 | 24.7   | 621 | 62 | AW440631    | AW440631 AT440631   |
| c          | 41    | 224.4 | 24.7   | 460 | 44 | AT795130    | AT795130 sbb77d1.y  |
|            | 42    | 218.4 | 24.0   | 720 | 54 | AT0108575   | AT0108575 cat060011 |
|            | 43    | 215.2 | 23.6   | 353 | 48 | AW058227    | AW058227 aa58702.y  |
|            | 44    | 214.6 | 23.6   | 378 | 40 | AW1460950   | AW1460950 aa7802.y  |
| c          | 45    | 212.8 | 23.4   | 616 | 47 | AT9399128   | AT9399128 70154548  |

## ALIGNMENTS

|            |                                                                 |        |      |     |             |
|------------|-----------------------------------------------------------------|--------|------|-----|-------------|
| RESULT     | 1                                                               |        |      |     |             |
| AW690945   |                                                                 |        |      |     |             |
| LOCUS      | AW690945                                                        | 662 bp | mRNA | EST | 17-APR-2000 |
| DEFINITION | NF034H11ST1F1000 Developing stem Medicago truncatula cDNA clone |        |      |     |             |
|            | NF034H11ST 5', mRNA sequence.                                   |        |      |     |             |

|           |                                                                                                                                                                      |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACCESSION | AW690945                                                                                                                                                             |
| VERSION   | AW690945.1 GI:7555504                                                                                                                                                |
| KEYWORDS  | EST.                                                                                                                                                                 |
| SOURCE    | barrel medic.                                                                                                                                                        |
| ORGANISM  | Medicago truncatula                                                                                                                                                  |
|           | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago. |
| REFERENCE | 1 (bases 1 to 652)                                                                                                                                                   |
| AUTHORS   | Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weiler,J.W. and May,G.D.                                                |
| TITLE     | Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research                                                             |
| JOURNAL   | Unpublished (2000)                                                                                                                                                   |
| COMMENT   | On Jan 6, 2000 this sequence version replaced gi:6676601.                                                                                                            |

## FEATURES

```
source      1. 662
            /organism="Medicago truncatula"
            /db_xref="taxon:3880"
            /clone="NF034111ST"
            /clone_lib="Developing stem"
            /tissue_type="stem"
            /dev_stage="Pooled developmental"
            /note="Vector: Lambda Zap; Contains a mixture of
            internal stem segments"
```

|            | Inter-nuclear beam segments |       |       |       |          |
|------------|-----------------------------|-------|-------|-------|----------|
| BASE COUNT | 207 a                       | 100 c | 140 g | 211 t | 4 others |
| ORIGIN     |                             |       |       |       |          |

Query Match 41.7%; Score 379.2; DB 79; Length 662;  
Best Local Similarity 81.9%; Pred. No. 1.3e-81;  
Matches 435: Conservative 0: Mismatches 96: Indels 0: Gaps 0:

Ov 2 aaaaacaatgagcactgcaagatttatcaagtgtgtcacggtcgggtgatggagctgtggg 61

Dh 121 AAGAAGAATGAGTACTGCTAGGTTTATCAAGTGTGTACAGTAGGGGATGGTGGTGTGG 180

On 62 gaaaattgtatgtcatttcatataccagcaatactttcccaacgattatgttccaac 121

Dh 181 AAAAACTTGCATGCTTATATCCTATACAAGCAATACCTTTCCCACTGATTATGTTCCAAC 240

On 122 agtattttgataattttatgtacacattatgtatgtatgagatggagagacagtgaacattgagct 181

Db 241 TCTGTTTCAGATTTCAGTCCGTAATCTAGCTGCGATGCTATACAGCTTAATCTTCGTTT 300

On 182 ntgggagcgtgcggcgccggacgctttatctactaggctacggcgagtgcatttatggcaga 241

201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

[illegible]

|   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 524 | 5 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|

||||||| | ||||| ||||| ||||| ||||| ||||| |||||

[illegible][illegible][illegible]



1

Db 223 TCGAAGTGAAGGCAACCTGTGCACCTCCACATCTGGGACACAGCAGGCGAAGTACT 282  
Qy 208 ataatagctgaagccactgagttatagagagctgattgttttttggcctttttc 267  
Db 283 ATGACGGCTGTGGCCCTGTCTACCTGACGCCGCTCTCTGCTTGTCTGATG 342  
Qy 268 ttataagcaggccagtttgaatacctacataaaagttagctccagagctaaagcatt 327  
Db 343 TCACGAGCCCAACAGCTTTGACAACTCTTTAACCGGTGTACCCAGAAGTGAATCATT 402  
Qy 328 atgtctaatgtaccagctgtgctgtttggaaccaaactagattgctgagatgacaa 385  
Db 403 TCTGCAAGAGGTACCATCATGCTGTGGGTGCAAGACTGACCTGTGCGCAAGGACAA 460

## RESULT 13

US-08-766-551-2

; Sequence 2, Application US/08766551

; Patent No. 5840569

## ; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Goli, Surya K.

; TITLE OF INVENTION: NOVEL HUMAN GTP-BINDING PROTEINS

; NUMBER OF SEQUENCES: 9

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/766,551

; FILING DATE: Herewith

; CLASSIFICATION: 514

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0168 US

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

## ; INFORMATION FOR SEQ ID NO: 2:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 719 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

## ; IMMEDIATE SOURCE:

; LIBRARY: SEQ ID NO:2

; CLONE: 113700

US-08-766-551-2

Query Match 10.24; Score 93; DB 3; Length 719;

Best Local Similarity 54.24; Pred. No. 4e-16;

Matches 189; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 52 gagctgtggggaacattgtatgctctattcatatcacagcaatactttcccaagggatt 111

Db 33 GCGGCGTTGGATGCAACACATTGGGGTGAGCTACACCCCAACGGCTTACCCACCGAGT 92

Qy 112 atgttccacagattttgataacttttagtgcacattgtgtgtgagtgagccacagtg 171

Db 93 ACATCCCTACTGCCTTCGACAACTCTCCGGGTGCTGCTGGATGGCGGCGCGTGA 152

Qy 172 aacttgacctatggacactgcggggaagagattatataagctaaagccactgatt 231

Db 153 GACTCCAACTCTGTGACACTGCCGACAGGATGAATTTGACAACTGAGGCTCTCTGCT 212

Qy 232 atagagagctgattgtttttgttgccttttcttataagcaggccagttatgaa 291

Db 213 ACACCAACACAGACATCTTCTGCTCTCTCAGTGTGTGAGCCCTCATCTCTCCAGA 272

Qy 292 acatctacaaaagtggatcccgagctaaagattatgtctataatgtaccagttgtgc 351

Db 273 AGTCACTGAGAAATGGGTGCGGAGATTGATGCCACTGTCCAAAGCCCCCATCTCC 332

Qy 352 ttgttggaacaaactgattgtgagatgacacagcagttctctattga 400

Db 333 TAGTGGAAACGACGTGCGATCTCAGAGAAGATGTCAAAGTCTCTCATTTGA 381

## RESULT 14

US-08-247-946-5

; Sequence 5, Application US/08247946A

; Patent No. 5792638

## ; GENERAL INFORMATION:

; APPLICANT: AARONSON, S.A.; CHAN, A.;

; APPLICANT: MIKI, T.

; TITLE OF INVENTION: NOVEL HUMAN RAS-RELATED

; TITLE OF INVENTION: ONCOGENES UNMASKED BY EXPRESSION OF cDNA

; TITLE OF INVENTION: CLONING

; NUMBER OF SEQUENCES: 11

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN &amp; FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/247,946A

; FILING DATE: 24-MAY-1994

; CLASSIFICATION: 536

## ; ATTORNEY/AGENT INFORMATION:

; NAME: DOROTHY R. AUTH

; REGISTRATION NUMBER: 36,434

; REFERENCE/DOCKET NUMBER: 2026-4150

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

## ; INFORMATION FOR SEQ ID NO: 5:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 615

; TYPE: Nucleic acid

; STRANDEDNESS: Double

; TOPOLOGY: Unknown

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: No

; ORIGINAL SOURCE:

; ORGANISM: Human

; STRAIN:

; INDIVIDUAL ISOLATE:

; DEVELOPMENTAL STAGE:

; HAPLOTYPE:

; TISSUE TYPE:

; CELL TYPE:

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Vers1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/842,976  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0267 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 702 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGNOT10  
CLONE: 1379718

Query Match 11.5%; Score 105; DB 5; Length 702;  
Best Local Similarity 55.7%; Pred. No. 2.5e-19;  
Matches 201; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 28 tcaaqtatgtcacqatcagatagaaactatagagaaaacttatatactcatttcata 87

Db 97 TCAAGGTGGTCTGGTGGGGGACGGCGGCTGCGGGAAGACGTGCTGCTGATGGTCTTCG 156

Ov 88 ccagcaatactttcccaacggattatgttccaacagtatttgcataacttttaagtcacaaatg 147

Db 157 CCGATGGGGCCCTCCCGGAGAGCTACACCCCCACGGTGTTTGAGCGGTACATGGTCAACC 216

Ov 148 tgcctggtgagatgacacacagctgaacctttgacctatgggacactgcccccaagaagatt 207

Db 217 TGCAGTGAAAGGCCAAACCTGTGCACCTCCACATGTGGGACACAGCAGGGCAGGAAGACT 276

Ov 208 ataataagcctaagccactgaattatagaggagctgatgtatttttattgacctttttctc 267

Db 277 ATGATCGACTGCGGCTCTCTCCTACCCGGACACTGATGTGCATCCTCATGTGCTTCTCCA 336

Ov 268 ttataagcaagggcagttatgaaaacatctacaaaaagtcgatcccagagctaagacatt 327

Db 337 TCGACAGCCCTGACAGCCTGGAAACATTTCCTGAGAAGTGGACCCCAGAGGTGAAGCACT 396

Ov 328 atgctcataatgtaccagttgtgcttgttgaaaccaaactagatttcgagatgacaagc 387

Db 397 TCTGCCCAACGTGCCCATCATCTGGTGGGGAATAAGAAGGACCTGAGGCAAGACGAGC 456

0v 388 a 388

Db 457 A 457

RESULT 10

US-09-213-397-2

; Sequence 2, Application US/09213397

; Patent No. 6063377

1. *Chlorophyll a* (Chl a) and *Chlorophyll b* (Chl b) are the primary photosynthetic pigments in green plants. They are responsible for capturing light energy and converting it into chemical energy through the process of photosynthesis. Chl a is the most abundant pigment, while Chl b is present in smaller amounts. Both pigments absorb light in the blue and red regions of the visible spectrum.

GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL HUMAN RHO PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/213,397  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/842,976  
FILING DATE: 04/17/1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0267 US  
TELECOMMUNICATIONS INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 702 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LONGMONT10  
CLONE: 1379718  
US-09-213-397-2.

Query Match 11.5%; Score 105; DB 5; Length 702;  
Best Local Similarity 55.7%; Pred. No. 2.5e-19;  
Matches 201: Conservative 0: Mismatches 160: Indels 0: Gaps 0:

Ov 28 tcaagtgtgtcacagtcagtgatggagctgtggggaaaactttgtatgctcatttcata 87

Db 97 TCAAGGTGGTCTCTGGTGGGCGACGGCGGCTGCGGGAAGACGTCGCTGCTGATGGTCTTTCG 156

Ov 88 ccagcaatactttcccaacggattatgttccaacagtatttgataactttaagtccaatg 147

db 157 CCGATGGGGCCTTCCCCGAGAGCTACACCCCCACGGTGTGTTGAGCGGTACATGGTCAACC 216

Ov 148 tggatggtggtatggcagcagcagtgaaaccttggcctatgggacactgcccggcaagaagatt 207

Db 217 TGCAGTGAAGGCCAAACCTGTGCACCTCCACATGTGGGACACAGCAGGGCAGGAAGACT 276

Ov 208 ataataggctaaagccactgaattatagaggagctgatatgtttttattgaacctttttctc 267

Db 277 ATGATCGACTGCGGCCTCTCTCCTACCCGGACACTGATGTTCATCCTCATGTGCTTCTCCA 336

Ov 268 ttataagcaagccagttatgaaaacatctacaaaaagtggatcccagagctaagacatt 327

Db 337 TCGACAGCCCTGACAGCCTGGAAACATTCTGAGAAGTGGACCCCAGAGGTGAAGCACT 396

Ov 328 atgctcataatgtaccagttgtcttctgttgaaccaaactagatttcgagatgacaac 387

$\frac{1}{2}$   $\frac{1}{4}$   $\frac{1}{8}$   $\frac{1}{16}$   $\frac{1}{32}$   $\frac{1}{64}$   $\frac{1}{128}$   $\frac{1}{256}$   $\frac{1}{512}$   $\frac{1}{1024}$   $\frac{1}{2048}$   $\frac{1}{4096}$   $\frac{1}{8192}$   $\frac{1}{16384}$   $\frac{1}{32768}$   $\frac{1}{65536}$   $\frac{1}{131072}$   $\frac{1}{262144}$   $\frac{1}{524288}$   $\frac{1}{1048576}$   $\frac{1}{2097152}$   $\frac{1}{4194304}$   $\frac{1}{8388608}$   $\frac{1}{16777216}$   $\frac{1}{33554432}$   $\frac{1}{67108864}$   $\frac{1}{134217728}$   $\frac{1}{268435456}$   $\frac{1}{536870912}$   $\frac{1}{1073741824}$   $\frac{1}{2147483648}$   $\frac{1}{4294967296}$   $\frac{1}{8589934592}$   $\frac{1}{17179869184}$   $\frac{1}{34359738368}$   $\frac{1}{68719476736}$   $\frac{1}{137438953472}$   $\frac{1}{274877906944}$   $\frac{1}{549755813888}$   $\frac{1}{1099511627776}$   $\frac{1}{2199023255552}$   $\frac{1}{4398046511104}$   $\frac{1}{8796093022208}$   $\frac{1}{17592186044416}$   $\frac{1}{35184372088832}$   $\frac{1}{70368744177664}$   $\frac{1}{140737488355328}$   $\frac{1}{281474976710656}$   $\frac{1}{562949953421312}$   $\frac{1}{1125899906842624}$   $\frac{1}{2251799813685248}$   $\frac{1}{4503599627370496}$   $\frac{1}{9007199254740992}$   $\frac{1}{18014398509481984}$   $\frac{1}{36028797018963968}$   $\frac{1}{72057594037927936}$   $\frac{1}{144115188075855872}$   $\frac{1}{288230376151711744}$   $\frac{1}{576460752303423488}$   $\frac{1}{1152921504606846976}$   $\frac{1}{2305843009213693952}$   $\frac{1}{4611686018427387904}$   $\frac{1}{9223372036854775808}$   $\frac{1}{18446744073709551616}$   $\frac{1}{36893488147419103232}$   $\frac{1}{73786976294838206464}$   $\frac{1}{147573952589676412928}$   $\frac{1}{295147905179352825856}$   $\frac{1}{590295810358705651712}$   $\frac{1}{1180591620717411303424}$   $\frac{1}{2361183241434822606848}$   $\frac{1}{4722366482869645213696}$   $\frac{1}{9444732965739290427392}$   $\frac{1}{18889465931478580854784}$   $\frac{1}{37778931862957161709568}$   $\frac{1}{75557863725914323419136}$   $\frac{1}{151115727451828646838272}$   $\frac{1}{302231454903657293676544}$   $\frac{1}{604462909807314587353088}$   $\frac{1}{1208925819614629174706176}$   $\frac{1}{2417851639229258349412352}$   $\frac{1}{4835703278458516698824704}$   $\frac{1}{9671406556917033397649408}$   $\frac{1}{19342813113834066795298816}$   $\frac{1}{38685626227668133590597632}$   $\frac{1}{77371252455336267181195264}$   $\frac{1}{154742504910672534362390528}$   $\frac{1}{309485009821345068724781056}$   $\frac{1}{618970019642690137449562112}$   $\frac{1}{1237940039285380274899124224}$   $\frac{1}{2475880078570760549798248448}$   $\frac{1}{4951760157141521099596496896}$   $\frac{1}{9903520314283042199192993792}$   $\frac{1}{19807040628566084398385987584}$   $\frac{1}{39614081257132168796771975168}$   $\frac{1}{79228162514264337593543950336}$   $\frac{1}{158456325028528675187087900672}$   $\frac{1}{316912650057057350374175801344}$   $\frac{1}{633825300114114700748351602688}$   $\frac{1}{1267650600228229401496703205376}$   $\frac{1}{2535301200456458802993406410752}$   $\frac{1}{5070602400912917605986812821504}$   $\frac{1}{10141204801825835211973625643008}$   $\frac{1}{20282409603651670423947251286016}$   $\frac{1}{40564819207303340847894502572032}$   $\frac{1}{81129638414606681695789005144064}$   $\frac{1}{162259276829213363391578010288128}$   $\frac{1}{324518553658426726783156020576256}$   $\frac{1}{649037107316853453566312041152512}$   $\frac{1}{1298074214633706907132624082305024}$   $\frac{1}{2596148429267413814265248164610048}$   $\frac{1}{5192296858534827628530496329220096}$   $\frac{1}{10384593717069655257060992658440192}$   $\frac{1}{20769187434139310514121985316880384}$   $\frac{1}{41538374868278621028243970633760768}$   $\frac{1}{83076749736557242056487941267521536}$   $\frac{1}{166153499473114484112975882535043072}$   $\frac{1}{332306998946228968225951765070086144}$   $\frac{1}{664613997892457936451903530140172288}$   $\frac{1}{1329227995784915872903807060280344576}$   $\frac{1}{2658455991569831745807614120560689152}$   $\frac{1}{5316911983139663491615228241121378304}$   $\frac{1}{10633823966279326983230456482242756608}$   $\frac{1}{21267647932558653966460912964485513216}$   $\frac{1}{42535295865117307932921825928971026432}$   $\frac{1}{85070591730234615865843651857942052864}$   $\frac{1}{170141183460469231731687303715884105728}$   $\frac{1}{340282366920938463463374607431768211456}$   $\frac{1}{680564733841876926926749214863536422912}$   $\frac{1}{1361129467683753853853498429727072845824}$   $\frac{1}{2722258935367507707706996859454145691648}$   $\frac{1}{5444517870735015415413993718908291383296}$   $\frac{1}{10889035741470030830827987437816582766592}$   $\frac{1}{217780$

Db 397 TCTGCCCAACGTGCCCATCATCCTGGTGGGAATAAGAAGGACCTGAGGCAAGACGAGC 456

Qy 388 a 388

RESULT 7  
 US-07-914-284A-6  
 : Sequence 6, Application US/07914284A  
 : Patent No. 5489524  
 : GENERAL INFORMATION:  
 : APPLICANT: Chow, Terry Y.-K.  
 : APPLICANT: Resnick, Michael A.  
 : APPLICANT: Perkins, Edward  
 : TITLE OF INVENTION: A CHIMERIC PROTEIN THAT HAS A HUMAN RHO  
 : TITLE OF INVENTION: MOTIF AND DIOXYRIBONUCLEASE ACTIVITY  
 : NUMBER OF SEQUENCES: 9  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Knobbe, Martens, Olson & Bear  
 : STREET: 620 Newport Center Drive, Sixteenth Floor  
 : CITY: Newport Beach  
 : STATE: CA  
 : COUNTRY: USA  
 : ZIP: 92660  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/07/914,284A  
 : FILING DATE: 14-JUL-1992  
 : CLASSIFICATION:  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/674,801  
 : FILING DATE: 26-MAR-1991  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Altman, Daniel P.  
 : REGISTRATION NUMBER: 34,115  
 : REFERENCE/DOCKET NUMBER: NIN022.022CPI  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (714) 760-0404  
 : TELEFAX: (714) 760-9502

Qy 28 tcaagctgtgtcaagctgtgtgtgagctgtggggaacactgtatgctcattcatata 87  
Db 471 TGAAGCTGTGTGGTGTGGGGGACGCTGCCCTGGGGAACCTGCCCTGATGAGCTACG 530  
Qy 88 ccagcactacttctccacgagattatgttccacagattattgtaacctttagtgcgaat 147  
Db 531 CCACACGAGCCTTCCGACGAGATACTGCTGCCACCTGTGTTGACCACTATGCATTTACTG 590  
Qy 148 tgggtgtggtatggcagacagactgaacctgtctatgggaacactgcgcggcaagaatt 207  
Db 591 TGACTCTGGGAGCAACACACACTTGTCTGGATGCTATGACACGCGGACAGAGAGACT 650  
Qy 208 ataatagcctaaagccactaattatacagagactatgtgttttggctcctttctctc 267

```
Db 92 TCAGGTGGTCCCTGGTGGGCGACGGCGGCTGCGGGAAGACCTGCTGCTGATGGGTCPTCG 151
Qy 88 ccagcactactttcccaacggtattgttcccaacgattttgataacttttagtgccaatg 147
Db 152 CCGATGGGGGCTTCCCGGAGAGCTACACCCCGAGGTGTTTGAGCGGTACATGGTCAAGC 211
Qy 148 tgggtggtgatggcagcacagtgaaaccttgccctatgggacactgcccggcagagagatt 207
Db 212 TGCAGGTGAAGGCAAGCAACCTGTGCACCTCCACATCTGGGACACAGCAGGGCAAGTGAAT 271
Qy 208 ataataaggctaaagccactgagttatagaggagctgatgtgttttggccttttctc 267
Db 272 ATGACCGGCTGCGGGCCCTGTCTACCCCTGACGCGCAGCGTCTGCTTGTGCTTCGATG 331
Qy 268 ttataagcaagccagttatgaaacatctacaaaaagtggatcccgagctaaagcatt 327
Db 332 TCACCGCCCGGCAACAGCTTTGACACACATCTTTAACCGGTGGTACCCGAGTGAATCAAT 391
Qy 328 atgctcataatgtaccagtttgctgtttggaacaaactagatttgagagatgacaa 385
Db 392 TCTGCAAGAGGTACCCATCATCTGTCGTGGGCTGCAAGACTGACCTGCGCAAGGACAA 449
```

Search completed: September 3, 2000, 03:10:10  
Job time: 28556 sec

dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, CC  
CC glomerulonephritis, gout, Graves' disease, hyperosmotic lability, irritable  
CC bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia  
CC gravis, myocardial or pericardial inflammation, osteoarthritis, CC  
CC osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, CC  
CC scleroderma, Sjogren's syndrome, and autoimmune thyroiditis. They can  
CC also be used to treat complications of cancer, hemodialysis, and  
CC extraocular circulation, viral, bacterial, fungal, parasitic,  
CC protozoal, and helminthic infections, and trauma. The products can also  
CC be used to treat diseases associated with apoptosis, such as AIDS and  
CC other infectious or genetic immunodeficiencies, neurodegenerative  
CC diseases such as Alzheimer's disease, amyotrophic lateral sclerosis,  
CC Parkinson's disease, retinitis pigmentosa and cerebellar degeneration,  
CC myelodysplastic syndromes such as aplastic anaemia, ischaemic injuries  
CC such as myocardial infarction, stroke and reperfusion injury, toxin-  
CC induced diseases such as cachexia, viral infections such as those caused  
CC by hepatitis B and C and osteoporosis.

Sequence 2966 BP: 822 A; 715 C; 641 G; 786 T;

Query Match 14.8%; Score 134.4; DB 1; Length 2966;  
Best Local Similarity 54.4%; Pred. No. 1.4e-24;  
Matches 295; Conservative 0; Mismatches 241; Indels 6; Gaps 1;

|    |     |                                                               |     |
|----|-----|---------------------------------------------------------------|-----|
| Qy | 28  | tcaagtggttcacggttcggtgtacagacttggggaacactgtatgctattcatata     | 87  |
| Db | 473 | TGAATGTGTGTGTGTGGGGGACGGTCCGCGGGGAAMCTGCCCTGCTAGTACGTACG      | 532 |
| Qy | 88  | ccagcgaactatttccacagcgattatgttccacagatttggataacttttagtgcgaagt | 147 |
| Db | 533 | CCACACGAGCGCTTCCGAGGAGATACGTGCCCMCTGTGTTGACCMCTATCAGTCTACTG   | 592 |
| Qy | 148 | tgggtgttgatggagcagcagtgaaacctggcctatgggaacctgcggcggaagaagatt  | 207 |
| Db | 593 | TGACTTGTGGGAGGACACACCMCTTGTGCGAGCTGTATGACACCGCGCAGGAGGACT     | 652 |
| Qy | 208 | ataataggctaaagccactgagttataggaagcgtctgtgtttttgtggtctttctc     | 267 |
| Db | 653 | ACACACGACTGAGGGCCMCTCTCTACCCACCGACGAGTGTGTTTGTATCTGCTCTCTG    | 712 |
| Qy | 268 | ttataagcaagcgagttatgaaacactctacaaagaattggtaccagcgaacagcatt    | 327 |
| Db | 713 | TGTAAACCTGCTCTTACACACATGCTCAGAGGAATGGTCCCGAGCTCAGGAAGCT       | 772 |
| Qy | 328 | atgtctataatgtacacagttgtgctttggacaacactagattcgogagatgc----     | 383 |
| Db | 773 | GCATGCTTCACTGCGCTTATGTGCTCTATGGAGACCCAGATGTGATCTCGTGATGACCCA  | 832 |
| Qy | 384 | ---aagcagttctctatgtatcacctctggagcaacacaaatatacaactctcagggaaga | 441 |
| Db | 833 | AAACCTTGGCCCTTCTCTGTATATGAAGAGAAACCTCTCATACGAGCATGGTGTA       | 892 |

Qy 442 aactaagaagatgatgatgagcagttacttattatagaatgcagctccaaaccccaacaga 501  
 Db 893 AGCTCGCAAAAGCAGTCTGGAGCACAGTGTCTACTGGAAATGTTCACTCTGACTCAAGAAG 952

Qy 502 atgtgaagctgttttgcagtgtgcataaagaatgctgtttaggcccaaaaacaaaga 561  
 Db 953 GTCTCAAAAGCGGTTTTGTGAAGCAATCCTCACACTTTTCCACCCCAAGAAAAAGA 1012

Qy 562 ga 563  
 |  
 Db 1013 AA 1014

```

RESULT 12
Q15017
ID   Q15017 standard; DNA; 2282 BP.
AC   Q15017;
DT   25-FEB-1992 (first entry)
DE   Encodes yeast endo-exonuclease RhoNUC.
KW   yeast cell cycle; rho/ras oncogene-like motif; RNC1 gene; ss.
OS   Saccharomyces cerevisiae.

```

```

PT Key Location/Qualifiers
FT cds 120..1577
FT /*tag= a
PN US7674801-A.
PD 05-MOV-1991.
PT 05-MOV-1991; 674801.
PR 26-MAR-1991; US-674801.
PA (US98 ) US DEPT HEALTH & HUMAN.
PI Resnick MA, Chow T, Perkins E;
PW WPI: 91-361692/49.
DR P-PDSB: RI5343.
PT Recombinant RhONC - useful for characterising agents to modify
PT cellular growth
PS Disclosure: Fig 2: 43pp; English.
CC The RNC1 gene was isolated from a yeast genomic library. It is
CC predicted to encode a protein of mol.wt. 57kD; the observed mol.wt.
CC is 72kD and the difference is thought to be due to glycosylation.
CC The N-terminal region of the deduced amino acid sequence shows
CC considerable homology with mammalian rho genes which are related to
CC ras oncogenes. The deduced C-terminal sequence has homology with
CC E.coli recC.
SQ Sequence 2282 BP: 770 A: 439 C: 435 G: 638 T:

```

Query Match 11.8%; Score 107.4; DB 1; Length 2282;  
Best Local Similarity 58.3%; Pred. No. 7e-18;  
Matches 208; Conservative 0; Mismatches 146; Indels 3; Gaps 1.

|    |     |                                                              |     |
|----|-----|--------------------------------------------------------------|-----|
| Qy | 28  | tcgaagtgtgcacggtcggtgtgagactgtgggaaactctatgctctattcatata     | 87  |
| Db | 334 | TAAAGATTGTTGTGTGGGAGATGGCGCTGTAGGAGAACGCTGCTTGATATTCTATG     | 393 |
| Qy | 88  | ccagcactattctccacacggattgattgtccaacgtatttgataactttagtcgcaagt | 147 |
| Db | 394 | TCCACAGAACATTCCGACGTATTATATTCCTACATTTTCGAAATATTGTCACAAACA    | 453 |
| Qy | 148 | tgggtgtggtgagcagacagcta--accttggcctattggacactgcggcgcaagaag   | 204 |
| Db | 454 | TAGAAGACCCACCGTCAAAATTAGAAITGCGATTGTGGACATCGCGGCCACGAAG      | 513 |
| Qy | 205 | attataataggctaaagcgaactgagttatagaggactgtagtctttttgtggccttt   | 264 |
| Db | 514 | AGTATAGTAGACTAGACCGCTTTTATATAGGAATCCAGATGTCGCTAGTGCGTATT     | 573 |
| Qy | 265 | ctcttatgaagcagcgactattgaacaactctacaanaagtcgacccaggaatgaagc   | 324 |
| Db | 575 | CTGTGTGCTTAGACATCGCTTAAAGATTGGAGATCTCTGTTGCCACAGGTTAAGC      | 634 |
| Qy | 324 | attattgctatcattatgacactgttctgttggagacacaaactgattgcgagatgt    | 381 |
| Db | 634 | ATTTTTGCTCTCCACCTCAATCATGCTATGCGGCTTAAATCAGATCATATGAAG       | 690 |

```

RESULT 13
V58232
ID V58232 standard; cDNA; 702 BP.
AC V58232;
DT 16-FEB-1999 (first entry)
DE Nucleotide sequence encoding human Rho
KW s; human; Rho protein; cell prolifer
KW translocation; cancer; gene therapy
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 45..663
FT /tsg= a
FT /product= 'Human Rho'
FT
PT
MN W090846754-1.
PD 22-OCT-1998
PR 16-APR-1998; W07865.
PR 17-APR-1997; U5-842976.
PA (JNCY) INCYTE PHARM INC.
PI Golli SK, Hillman JL;
DI WPI: 98-609916/51.

```

DE Candida CaCd42 gene.  
 KW GPase; GGPase; geranylgeranyl transferase; fungal Rho-like GPase;  
 antifungal agent; identification; mycosis; feed additive; disinfectant;  
 therapy; Candida cell detection; cell wall integrity; hyphal formation;  
 pathogenesis; Candida CaCd42 gene; ds.  
 OS Candida sp.  
 FH Key Location/Qualifiers  
 FT CDS 260..835  
 FT /\*tag= a  
 PN W09738293-A2.  
 DE 16-OCT-1997.  
 PF 11-APR-1997; U05987.  
 PR 20-DEC-1996; US-771212.  
 PR 11-APR-1996; US-631319.  
 PA (MITO-) MITOFIX INC.  
 PA (UTJO) UNIV JOHNS HOPKINS.  
 DR WPI: 97-512864/47.  
 DR P-PSDB: W33897.  
 PT Identification of antifungal agents that inhibit GPase - useful for  
 PT specific detection of Candida  
 PS Claim 115; Page 84-85; 118pp; English.  
 CC This sequence represents the Candida CaCd42 gene. The encoded protein is  
 CC a fungal Rho-like GPase. The encoded protein can be used in an assay of  
 CC the invention. The method of the invention is for identifying potential  
 CC antifungal agents (I), and comprises: (a) mixing a fungal geranylgeranyl  
 CC transferase (GGPase), a GGPase substrate GGPase, and test compound;  
 CC and (b) detecting interaction between GGPase and GGPase; a significant  
 CC reduction in this interaction indicates that the test compound is a (I).  
 CC (I) are useful for treating mycoses in humans or animals; as feed  
 CC additives and as disinfectants. This sequence, and Mdb specifically  
 CC reactive with the encoded protein are used to detect Candida cells  
 CC specifically, particularly in cells, tissues and body fluids, while  
 CC antisense sequences are used to inhibit expression of these genes. The  
 CC method is a rapid, reliable and effective way of detecting agents that  
 CC inhibit GPases, particularly those involved in cell wall integrity,  
 CC formation of hyphae and/or other cellular functions necessary for  
 CC pathogenesis. (I) should be selective for fungal cells, with little  
 CC effect on mammalian cells.  
 SQ Sequence 934 BP; 290 A; 152 C; 157 G; 335 T;

Query Match 16.4%; Score 149.2; DB 1; Length 934;  
 Best Local Similarity 56.1%; Pred. No. 2e-28;  
 Matches 305; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

Yy 26 tatcaagtggtgcacggtggtgatggagctgtgggaacattgtatgctcatccta 85  
 Db 268 TATAAATGCTGTTGTTGTCGGTGATGGTGGCTGTGTAACCTGCTTATTAATCTGTA 327

Yy 86 taccagcactactttcccaacggtattgttccacagatttgataactttagtgcaca 145  
 Db 328 TACCACTAGTAATTTCCAGCTGATTATGTCTCACTGTTTGTATTAATTCGCTTAC 387

Yy 146 ttgtgtgtgtgatggcagcagctgaaccttgccatttggaacactggcgggcaagaaga 205  
 Db 388 CGTAGATAGAGAGCAACACCTTACCTTGGGATATTGTGATCTGCTGGTCAAGAGA 447

Yy 206 ttataataggtaggacccattgattatagaggagctgattgtttttgtgcccatttc 265  
 Db 448 TTACAGCAGATTAAGGCCCTTTGTCATATCCATCGACTGATTAATCTGTTTGTGTTTC 507

Yy 266 tcttataagcagcgccattgtgaaacactctcaaaaagtggatccagagcttaagaca 325  
 Db 508 CCGTATTCTCTGCTGCTGTTGTAATGTTAAAGAAAATGGTCCCAAGAGTTCATCA 567

Yy 326 ttatgctcataatgtaccagttgtgtgtgttggaacaaactagattgtgcagatgaca 385  
 Db 568 CCAATTGCTGCGGTGTGCAATAATTAATGTGCGGTACCCAACTGATTTAGAAAGCATGA 627

Yy 386 gcaatttctattgatccacccctggagaca -----taccacatacacaactctcaaggaga 439  
 Db 628 TGTTATTTTACAGAGATTGACAGACAAAATGTGCCCAATCACCCAGGAACGGTGA 687

Yy 440 agaactaaagagatgataggagcgtacttatatagaatgcagctccaaaaccaca 499  
 Db 688 AAATTCGCTAGAGATTGAGAGCTGTCAAGTATGTGTAGTGTCTGATTCATCAAG 747

Yy 500 gaatgtgaagctgttttgcagtctgcaataaagtagctttggagccacaaaaccaca 559  
 Db 748 AGGATTGAAAAACGCTTTGACAGCGTATAGTGCCTAGTAACCTCTGTAATTA 807

Yy 560 gaga 563  
 Db 808 AAAA 811

RESULT 9  
 V32555  
 ID V32555 standard; RNA; 3243 BP.  
 AC V32555;  
 DT 13-OCT-1998 (first entry)  
 DE Candida albicans CaCd42p gene.  
 KW CaCd42p; G-protein; rho factor; screening; virulence;  
 KW hyphal formation; pathogenic fungi; inhibitor; inflammation;  
 KW antimycotic; ss.  
 OS Candida albicans.  
 FH Key Location/Qualifiers  
 FT CDS 271..846  
 FT /\*tag= a  
 FT /product= CaCd42p protein  
 PN W09818927-A1.  
 DR 07-MAY-1998.  
 PF 29-OCT-1997; CA0809.  
 PR 30-OCT-1996; US-029458.  
 PA (CAN) NAT RES COUNCIL CANADA.  
 PI Leberer E, Thomas D;  
 DR WPI: 98-27222/24.  
 DR P-PSDB: W48897.  
 PT In vitro screening test for agents that inhibit Candida genes  
 PT involved in virulence - and transition to hyphal form, potentially  
 PT useful as antimycotic agents  
 PS Disclosure; Fig 1; 79pp; English.  
 CC The sequence is that encoding the CaCd42p protein which can be used  
 CC in the development of an in vitro screening test for compounds  
 CC that inhibit biological activity of the protein and a system for  
 CC measuring its activity. The protein is involved in virulence and  
 CC hyphal formation. Inhibitors are potentially useful for rendering  
 CC pathogenic fungi (any species in which hyphal induction by kinase  
 CC occurs) avirulent and/or to treat inflammation. The coding sequence  
 CC can be used as source of probes for detecting C. albicans in  
 CC amplification or hybridisation assays, also to identify and  
 CC clone homologous genes from other fungi.  
 SQ Sequence 3243 BP; 1185 A; 541 C; 456 G; 1061 T;

Query Match 16.4%; Score 149.2; DB 1; Length 3243;  
 Best Local Similarity 56.1%; Pred. No. 3e-28;  
 Matches 305; Conservative 0; Mismatches 233; Indels 6; Gaps 1;

Yy 26 tatcaagtggtgcacggtggtgatggagctgtgggaacattgtatgctcatccta 85  
 Db 279 TATAAATGCTGTTGTTGTCGGTGATGGTGGCTGTGTAACCTGCTTATTAATCTGTA 338

Yy 86 taccagcactactttcccaacggtattgttccacagatttgataactttagtgcaca 145  
 Db 339 TACCACTAGTAATTTCCAGCTGATTATGTCTCACTGTTTGTATTAATTCGCTTAC 398

Yy 146 ttgtgtgtgtgatggcagcagctgaaccttgccatttggaacactggcgggcaagaaga 205  
 Db 399 CGTAGATAGAGAGCAACACCTTACCTTGGGATATTGTGATCTGCTGGTCAAGAGA 458

Yy 206 ttataataggtaggacccattgattatagaggagctgattgtttttgtgcccatttc 265  
 Db 459 TTACAGCAGATTAAGGCCCTTTGTCATATCCATCGACTGATTAATCTGTTTGTGTTTC 518

Yy 266 tcttataagcagcgccattgtgaaacactctcaaaaagtggatccagagcttaagaca 325



Qy 430 ctcagggaagaactaaagaagatgataggagcagttacttatagatgcagctcca 489  
 Db 550 CTGAAGGCCACAGTGTCTCAAGAATGTGTCTGCTGATTCTGGAAATGTTCTGCTA 609  
 Qy 490 aaaccacacagaatgtgaagcgtgttttcgtctgctgaataaagatgctttgagccac 549  
 Db 610 AACCGGTAGAGTGTGTAGAGAAGTGTTTGAAGCTGCTAGTAGCTCTTTTAAGAGTTA 669  
 Qy 550 caaacacacagaagaag 566  
 Db 670 AAGAAAAGAGGAAAG 686

RESULT 5  
 T92702  
 ID T92702 standard; cDNA; 3198 BP.  
 AC T92702;  
 DT 30-APR-1998 (first entry)  
 DE Candida Carhol gene.  
 KW GPase; GGTPase; geranylgeranyl transferase; fungal Rho-like GPase;  
 KW antifungal agent identification; mycosis; feed additive; disinfectant;  
 KW therapy; Candida cell detection; cell wall integrity; hyphal formation;  
 KW pathogenesis; Candida Carhol gene; ds.  
 OS Candida sp.  
 FH Key Location/Qualifiers  
 FT CDS 1362..1959  
 FT /\*tag= a  
 PN W09738129-A1.  
 PD 16-OCT-1997.  
 PF 10-APR-1997; U05929.  
 PR 10-APR-1996; US-631319.  
 PA (MITO-) MITOTIX INC.  
 PA (UOYO) UNIV JOHNS HOPKINS.  
 PI Berlin V, Damagomez V, Smith SE;  
 DR WPI: 97-512735/47.  
 DR P-PSDB: W30379.  
 PT Identification of antifungal agents that inhibit GPase - useful for  
 PT specific detection of Candida  
 PS Claim 118; Fig 16; 123pp; English.  
 CC This sequence represents the Candida Carhol gene. The encoded protein is  
 CC a fungal Rho-like GPase. The encoded protein can be used in an assay of  
 CC the invention. The method of the invention is for identifying potential  
 CC antifungal agents (1), and comprises: (a) mixing a fungal geranylgeranyl  
 CC transferase (GGTPase), a GGTPase substrate (A), and test compound;  
 CC and (b) detecting interaction between GGTPase and (A); a significant  
 CC reduction in this interaction indicates that the test compound is a (1).  
 CC (1) are useful for treating mycoses in humans or animals; as feed  
 CC additives and as disinfectants. This sequence, and WAB specifically  
 CC reactive with the encoded protein are used to detect Candida cells  
 CC specifically, particularly in cells, tissues and body fluids, while  
 CC antisense sequences are used to inhibit expression of these genes. The  
 CC method is a rapid, reliable and effective way of detecting agents that  
 CC inhibit GPases, particularly those involved in cell wall integrity,  
 CC formation of hyphae and/or other cellular functions necessary for  
 CC pathogenesis. (1) should be selective for fungal cells, with little  
 CC effect on mammalian cells.  
 SQ Sequence 3198 BP; 1057 A; 538 C; 517 G; 1086 T;

Query Match 18.04; Score 163.8; DB 1; Length 3198;  
 Best Local Similarity 57.34; Pred. No. 6.8e-32;  
 Matches 319; Conservative 0; Mismatches 232; Indels 6; Gaps 1;

Qy 16 ctcagagattatacagtgctgcacgggtggtgagcagctgtgggaacattgtatgc 75  
 Db 1378 CTGAAGGCCACAGTGTCTCAAGAATGTGTCTGCTGATTCTGGAAATGTTCTGCTA 1437  
 Qy 76 tcaattcatatcacagcactacttcccaagcgtatgtgtccacagctatttgataact 135  
 Db 1438 TAATTGTTTCTCAAGAGGTACTTTCCAGAGTTTATGCTCCACAGTGTCTTGAAGTT 1497  
 Qy 136 ttatgtccaaattgtggtgtggtgagcagcagtgaaacttggtcattggacactgccc 195

Db 1498 ACCTGCTGATGTGTGAAGTGTATGTTAGAAAAGTGAATGGCATATGGGACTACTGCTG 1557  
 Qy 196 ggcagagagattataataggcagcagcagctagttatagaggagcgtgtgtttttgt 255  
 Db 1558 GTCAAGAGATATGATAGATTAGACCAATCTCTTATCCAGATTCTAAGTTATTGGA 1617  
 Qy 256 tggcctttttcttataatagcagcagcagttatgaaacatctacaaaagtggatccag 315  
 Db 1618 TTGCTTTTTCAGTGAATCCACCAATCTTTAGATAAGCTTTAGAAAATGAGTTCTG 1677  
 Qy 316 agctaacacatttctctcattatgaccagttgtgctgttggaacacaaatagattgc 375  
 Db 1678 AAGTTTATCTTCTGCTCAAGGTGTTCACATCATTTAGTGTGTATAATCTGATTAA 1737  
 Qy 376 gagatgacagacagcttctct-----cattgatcccttgagacacacacatatcaact 429  
 Db 1738 GAGATGATCTCTACTACTTGAAGCTTGAGACACACACACACACCACTTCAACTT 1797  
 Qy 430 ctcagggaagaactaaagaagatgataggagcagttacttatagatgcagctcca 489  
 Db 1798 CTGAAGGCCACAGTGTCTCAAGAATGTGTCTGCTGATTCTGGAATGTTCTGCTG 1857  
 Qy 490 aaaccacacagaatgtgaagcgtgttttcgtctgctgaataaagatgctttgagccac 549  
 Db 1858 AACCGGTAGAGTGTGTAGAGAAGTGTTTGAAGCTGCTAGTAGCTCTTTTAAGAGTTA 1917  
 Qy 550 caaacacacagaagaag 566  
 Db 1918 AAGAAAAGAGGAAAG 1934

RESULT 6  
 T92869  
 ID T92869 standard; cDNA; 3198 BP.  
 AC T92869;  
 DT 30-APR-1998 (first entry)  
 DE Candida Carhol gene.  
 KW GPase; GGTPase; geranylgeranyl transferase; fungal Rho-like GPase;  
 KW antifungal agent identification; mycosis; feed additive; disinfectant;  
 KW therapy; Candida cell detection; cell wall integrity; hyphal formation;  
 KW pathogenesis; Candida Carhol gene; ds.  
 OS Candida sp.  
 FH Key Location/Qualifiers  
 FT CDS 1362..1959  
 FT /\*tag= a  
 PN W09738293-A2.  
 PD 16-OCT-1997.  
 PF 11-APR-1997; U05987.  
 PR 20-DEC-1996; US-771212.  
 PR 11-APR-1996; US-631319.  
 PA (MITO-) MITOTIX INC.  
 PA (UOYO) UNIV JOHNS HOPKINS.  
 DR WPI: 97-512864/47.  
 DR P-PSDB: W33896.  
 PT Identification of antifungal agents that inhibit GPase - useful for  
 PT specific detection of Candida  
 PS Claim 118; Fig 16; 118pp; English.  
 CC This sequence represents the Candida Carhol gene. The encoded protein is  
 CC a fungal Rho-like GPase. The encoded protein can be used in an assay of  
 CC the invention. The method of the invention is for identifying potential  
 CC antifungal agents (1), and comprises: (a) mixing a fungal geranylgeranyl  
 CC transferase (GGTPase), a GGTPase substrate GGTPase, and test compound;  
 CC and (b) detecting interaction between GGTPase and GGTPase; a significant  
 CC reduction in this interaction indicates that the test compound is a (1).  
 CC (1) are useful for treating mycoses in humans or animals; as feed  
 CC additives and as disinfectants. This sequence, and WAB specifically  
 CC reactive with the encoded protein are used to detect Candida cells  
 CC specifically, particularly in cells, tissues and body fluids, while  
 CC antisense sequences are used to inhibit expression of these genes. The  
 CC method is a rapid, reliable and effective way of detecting agents that  
 CC inhibit GPases, particularly those involved in cell wall integrity,  
 CC formation of hyphae and/or other cellular functions necessary for

Db 121 CAGTATTGTGAACCTTTAGTGCACATGTGTGTGTGAGTGACACAGCTGACACTTGTGCC 180

Qy 181 tatgtggaactgcgcgcgaagaattataatagctgaagccactgattatagaagag 240

Db 181 TATGGGACATGCTCGGGCAGAGAATTATATAGCTTAGGGCCATGAGTTATAGAGAG 240

Qy 241 ctgagtgttttttggctgtttcttcttataagcaagcgcagctatgaaaacatctaca 300

Db 241 CTGAGTGTGTTTGTGTGCTTTTCTCTATAGCAGGCGAGCTATGAAAACATCTACA 300

Qy 301 aaaaagtgtccagagctgaagcattatgctataatgacacagctgtgtgtgtgaa 360

Db 301 AAAAGTGTGTTCCGAGCTAAGACATTATGCTCATTAATGTACCACTGTGTGTGTGGA 360

Qy 361 ccaactgagtttcggagtgacagcagcttctctatgatccctgtgagcaacccaa 420

Db 361 CCAACTGATTTGCGAGATGACAAGCAGTCTCTCATGTATGACCTGTGAGCAACACCA 420

Qy 421 tatcacactctcagggaagaactaaagaagatgatagagcagttactattatagaat 480

Db 421 TATCACACTATCAGGGAAGAAGCTAAGAAGATGATAGAGCGTACTTATATAGAA 480

Qy 481 gcagctcccaaaacccaaagagtggaagctgttttggatgctgcaataaagtagctt 540

Db 481 GCGAGCTCCAAACCCAAAGAAATGTGAAGGCTGTTTGTGATGTCGAATAAAGTACTT 540

Qy 541 tgaggccaccacaacaaagagaagcgttgcaaaagagagaacatgtgttcttctgaa 600

Db 541 TGAGGCTCCAAACCCAAAGAAAGGCTTCCAAAGGAGACATGTGCTTCTTTGAA 600

Qy 601 tattgatcattattacagcaaaaacagtttaacaaaagctgtgcagetaaacactgaa 660

Db 601 TATTGATCATTAATTACAGCAAAAACAGTTAACAAAAGCTGTGCAGATAAACACTGAA 660

Qy 661 tctgtatagtttggtttggtttacctatgttccacgtgaaactatgaagcatctcaa 720

Db 661 TCTGTATAGTTTGGTTTGGTTTACATATGTTCACAGTGAAATGATAGAGCATCTCAA 720

Qy 721 gaaaacccaaactatcatcaacccctgatcatgaatgatatttcttctgagta 780

Db 721 GAAAACCCAAACTATCATATCAACCCATGATCAATGAATGATTTCAATTTTGCAGTA 780

Qy 781 taagtctttaaactcttcttcttcttcttcttcttcttcttcttcttcttcttctt 840

Db 781 TAAGTCTCTTTAACTCTCTCTTTTACTCATTTTATAAGCAATCTATGATATGTT 840

Qy 841 ccttcaaacatgttactacagttttatataaaatccattcttcttcttcttcttctt 900

Db 841 CCTTCAACATGTCATACAGTGTAAATATAAATTCATCTCTTCTATTTACTAAAA 900

Qy 901 aaaaaaaaaa 910

Db 901 AAAAAAAAAA 910

## RESULT 2

T73866  
ID T73866 standard; DNA; 3045 BP.  
AC T73866;  
DT 26-JAN-1998 (first entry)  
DE Cotton fibre promoter clone Rac13 construct, pCGW4735.  
KW promoter; fibre-specific; transcriptional factor; promoter;  
KW altered phenotype; colour; melanin; indigo; ss.  
OS Gossypium hirsutum cv. coker 130.  
PN W09640924-A2.  
PD 19-DEC-1996.  
PF 07-JUN-1996; U09897.  
PR 07-JUN-1995; US-480178.  
PR 01-JUL-1996; ZA-005572.  
PA (CALJ) CALGENE INC.  
PI McBride K, Pear JR, Perez-Grau L, Stalker DM;  
DR WPI; 97-052325/05.

PT DNA construct contg. gene of interest controlled by cotton fibre  
PT transcriptional factor - used to produce altered phenotype cotton  
PT fibre cells expressing genes affecting pigmentation  
PS Claim 23; Fig 5A-B; 8pp; English.  
CC The present sequence is the Rac13 promoter construct, pCGW4735, isolated  
CC from cotton fibre genomic clone 15-1. DNA constructs containing  
CC cotton fibre-specific transcriptional factor promoters are useful to  
CC produce cotton fibre cells with altered phenotype, especially altered  
CC colour. Genes involved in the production of melanin (e.g. tyrosinase  
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and  
CC indigo (mono-oxygenase genes possibly in conjunction with a  
CC tryptophanase gene) are of interest. The promoters of the invention are  
CC reliable and permit expression of a protein selectively in cotton fibre  
CC to affect qualities such as fibre strength, length, colour and dyability  
CC as required. The construct and methods can also be used for the  
CC introduction of other advantageous genes into a cotton plant, e.g. a  
CC plant hormone. In particular, fibres from a plant producing coloured  
CC fibres may be used to produce yarns and/or fabrics that do not require  
CC dyeing.

SQ Sequence 3045 BP; 1063 A; 450 C; 366 G; 1162 T;

Query Match 33.54; Score 304.8; DB 1; Length 3045;  
Best Local Similarity 97.84; Pred. No. 4.6e-67;  
Matches 309; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 594 ctttgaatttgatcatattatcagtcacaaacagcttaacaaagctgtgagataaa 653

Db 1805 CTCCTGAATTTGGATCATATTATCAGCTCAAAACAGCTTAAACAAAGCTGTTGAGATAAA 1865

Qy 654 caatgtaatgtctatagttgttttggtttatactatgttccagctgaactatgaaagca 713

Db 1856 CACTGAATCTGCTATAGTTGTTTGGTTTACATATGTTTCCAGCTGAAACTGATGAAGCA 1925

Qy 714 tctctagaagaaccccaactatcatatcaacccctgatcatgaatgatatttcttctt 773

Db 1926 TCTCTAGAAGAAACCCAACTATCATATCAACCTGATCATGAATGATGTTCAATTTT 1985

Qy 774 cgcagataagtttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 833

Db 1986 CGCAGTATAAGTCTCTTTTAACTCTCTTTTACTTCATTTTATAACGAATTCATGGA 2045

Qy 834 taatgttccctcaaacatgctatatacagtttaattataaaatccattcttcttctt 893

Db 2046 TAATGTTCCCTCAAAACATGTCATTACATGTTTAAATATAAATTCATCTCTTCTATTT 2105

Qy 894 actaaaaa 909

Db 2106 ACTAAGATATTAGTAA 2121

## RESULT 3

T92698  
ID T92698 standard; cDNA; 985 BP.  
AC T92698;  
DT 30-APR-1998 (first entry)  
DE Candida Carhol gene.  
KW GTPase; GGPase; geranylgeranyl transferase; fungal Rho-like GTPase;  
KW antifungal agent identification; mycosis; feed additive; disinfectant;  
KW therapy; Candida cell detection; cell wall integrity; hyphal formation;  
KW pathogenesis; Candida Carhol gene; ds.  
OS Candida sp.  
FH Key Location/Qualifiers  
FT CDS 114..710  
FT /\*tag= a  
PN W09738129-AL  
PD 16-OCT-1997.  
PF 10-APR-1997; U05929.  
PR 10-APR-1996; GS-631319.  
PA (MITO-) MITOXIL INC.  
PA (UKYO) UNIV JOHNS HOPKINS.  
PI Berlin V, Damaguer V, Smith SE;  
DR WPI; 97-512735/47.

Db 675 CAAACCCAGGGGAGAGGAACCTGAAGAACTGATCGATCTGCTCTACATTGAATGTAGTT 734  
Qy 487 ccaaaacccaacagaatgtgaaggctgttttgcgatgctgcataaaagttagcttggagc 546  
Db 735 CAAAGACACAGCGAAGCTGAAGGCACTGTTTGTGCGAGCTATAAAGTGGTGCTTCAGC 794  
Qy 547 caccnaaaccaagagaagccttgcaaaaggagaacatgtgctttccttga 599  
Db 795 CACCAAGCGAAGAGAAGAAAAGAAATAAGAACCCGTCGCGCTCTCTGTGA 847

Search completed: September 3, 2000, 04:03:47  
Job time: 35777 sec

Db 826 ATCCCTGAGCTCAGACACTATCGCCATCGGTACCATCATCTCTGCTGGGAGAGCTA 885  
 Qy 369 gatttggagatgacagacagttcttcttattgataccctggagcaacacaaatcaaca 428  
 Db 886 GATCTTCAGATGATTAACGATCTCTTCTGACCAATCTCGGAGCGGCTCCAAATACACC 945  
 Qy 429 tctcaggaggaagactaaagaatgatagagacagttcttattatagatgacgtcc 488  
 Db 946 TCTCAGGCGAGAGCTCAGGAAGTCAATTTGAGCGGCTTGATATTTGATGTACTCG 1005  
 Qy 489 aaaccccaacagaatggaagcgtgtttgatgctgcaataaagtagctttggagcca 548  
 Db 1006 AAGACTCAACAGATGTGAAGCAGTTTTGTATGACGACATCAAGGTGGTCTTCAGCCA 1065  
 Qy 549 ccaaaaccaagaagaag 566  
 Db 1066 CCCAAGCAGAAGAAG 1083

RESULT 13  
 AF115476  
 LOCUS AF115476 1558 bp mRNA PLN 20-APR-1999  
 DEFINITION Physcomitrella patens rac-like GTP binding protein (rac2) mRNA, complete cds.  
 ACCESSION AF115476  
 VERSION AF115476.1 GI:4588757  
 KEYWORDS  
 SOURCE Physcomitrella patens.  
 ORGANISM Physcomitrella patens  
 Buxariota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida; Bryidae; Funariales; Funariaceae; Physcomitrella.  
 REFERENCE 1 (bases 1 to 1558)  
 AUTHORS Winge, P., Kristensen, R., Bones, A.M. and Reski, R.  
 TITLE The Physcomitrella patens rac-gene family  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1558)  
 AUTHORS Kristensen, R., Winge, P., Bones, A.M. and Reski, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-1998) The Norwegian University of Science and Technology, UNIGEN MTFS, Olav Kyrresgate 3, Trondheim N7005, Norway

FEATURES  
 source Location/Qualifiers  
 1. 1558  
 /organism="Physcomitrella patens"  
 /db\_xref="taxon:3218"  
 gene 1. 1558  
 /gene="rac2"  
 5'UTR 1. 530  
 /gene="rac2"  
 CDS 531..1121  
 /gene="rac2"  
 /note="PhRac2"  
 /codon\_start=1  
 /product="rac-like GTP binding protein"  
 /protein\_id="AAD26198.1"  
 /db\_xref="GI:4588758"  
 /translation="MSTSRPFIKIVTGDGAVGKCOLLSTINFTDPTIVTDFDNFS  
 ANVVYGVTVNLGLNDAGQEDYNRLAPLSTRGADVPLLAFLSLKASTENISKKWIP  
 ELRHAPSPVILVLTGLDLRDKQFPADHGAARITTTSGEELRKSIGAASTIECSS  
 KTKQWNAVFDAKVKVLPQPKRKKKKKKQKCVIL"  
 1122..1558  
 /gene="rac2"  
 3'UTR  
 BASE COUNT 368 a 309 c 427 g 454 t  
 ORIGIN

Query Match 37.9%; Score 344.8; DB 8; Length 1558;  
 Best Local Similarity 72.7%; Pred. No. 5.6e-61;  
 Matches 445; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 5 aacaatgacatgcagatttatcaagtgtctcaagctgctgtagacgtgtgagaa 64  
 Db 527 AGCCATGACACTTCACGGTTATCAAGTCGGTCACTGTGGAGATGAGACTGTCCGGA 586

Qy 65 aactgttatgtctatttatacagcagatactttccacagattatgttcaacagt 124  
 Db 587 GAGCTGCATCTTATTTATACACAGCAGCACTTCTACTGATAGCTTCTACCGT 646  
 Qy 125 attttgataacttattgccaatgttggttggttgatggacacagtgtaacttggctatg 184  
 Db 647 GTTTCACACTCTCAGCGAAATGTAGGGTCGATGAATACCTGCACCTCGGGTTATG 706  
 Qy 185 ggcactctggggcagaagaattataatagctaaagccatgattatagagagctga 244  
 Db 707 GSATACAGCAGTCAAGAAGATTAACACAGCTGTCTCTCTGAGTTACAGGGGTCTGA 766  
 Qy 245 tgtgtttttgtgcttttctcttataagcaggccagttatgaacatctcaaaaa 304  
 Db 767 TGTCTTTCTCTGGGCTTCTCCCTCATCAGCAGGCTAGTATGAACATATCAAGAA 826  
 Qy 305 gtggatccagagatgaacattatgtctataattacacattgtgtgttggaacaa 364  
 Db 827 GTGGATCCGGAATGAGACATACGGCCATCTGTGCAATCATCTCTGTGGACAAA 886  
 Qy 365 actagatttgcagatgacacagcttctctattgacacccctggagcaacacaaatc 424  
 Db 887 ACTTGATCTTCGGATGACAACAATCTTTGCTCATCTCTGGAGCGCTCCATAC 946  
 Qy 425 aacatctcaggagagaaactaaagaatgatagagacattacttatatagatgcag 484  
 Db 947 TACTTTTCAGGGGAGAGCTCAGGAATGATTTGGGGCGGCTCTATAGATGACAG 1006  
 Qy 485 ctccaaaccccaagaatgtgaaggtgttttgcagatgcgaataaagtagctttgag 544  
 Db 1007 CTCAGAGCTCAGCAGATGTAAGAGCAGTTTTTGACACGACATCAGGTGTGTCCA 1066  
 Qy 545 gcaacaaacaaagaagaagccttgcgaagaagaatgtgttcttcttgatatt 604  
 Db 1067 ACCACCAAGCAGAAGAAGAAGAAAACAAAGAATTCGCTCATCTTGTGAATGTG 1126  
 Qy 605 ggatcattatta 616  
 Db 1127 GCATAGCTTTTA 1138

RESULT 14  
 NTA250174  
 LOCUS NTA250174 803 bp mRNA PLN 06-OCT-1999  
 DEFINITION Nicotiana tabacum mRNA for putative rac protein (rac gene).  
 ACCESSION AJ250174  
 VERSION AJ250174.1 GI:6015626  
 KEYWORDS rac gene; rac protein.  
 SOURCE common tobacco.  
 ORGANISM Nicotiana tabacum  
 Buxariota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 REFERENCE 1 (bases 1 to 803)  
 AUTHORS Kieffer, F., Elmayan, T., Simon-Plas, F., Dagher, M.C. and Blein, J.P.  
 TITLE A tobacco cDNA encoding a Rac-like protein cloned using the two-hybrid system in an heterologous screen  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 803)  
 AUTHORS Elmayan, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-OCT-1999) Elmayan T., UMR 692 INRA/Universite de Bourgogne, INRA / laboratoire de Phytopharmacie, BP 1540, 21034 Dijon Cedex, FRANCE

FEATURES  
 source Location/Qualifiers  
 1. 803  
 /organism="Nicotiana tabacum"  
 /cultivar="Xanthi"  
 /db\_xref="taxon:4097"  
 /dev\_stage="55 day-old plants"  
 /tissue\_type="young leaves: the nearest from apex"  
 /clone="Rac5"

```

RESULT 10
ATU64919
LOCUS
    ATU64919      1008 bp      mRNA           PLN      05-JAN-1999
DEFINITION
    Arabidopsis thaliana geranylgeranylated protein ATGP2 mRNA,
    complete cds.
ACCESSION
    U64919
VERSION
    U64919.1  GI:4097562
KEYWORDS
SOURCE
    Arabidopsis thaliana
    Eukarya: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons: core
    eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
    Arabidopsis.
REFERENCE
    1 (bases 1 to 1008)
AUTHORS
    Biermann,S.J., Price,J.R., Crowell,D.N. and Randall,S.K.
TITLE
    A collection of cDNAs encoding isoprenylated plant proteins
JOURNAL
    Unpublished
REFERENCE
    2 (bases 1 to 1008)
AUTHORS
    Biermann,S.J., Price,J.R., Crowell,D.N. and Randall,S.K.
TITLE
    Direct Submission
JOURNAL
    Submitted (23-JUL-1996) Biology, IUPUI, 723 West Michigan Street,
    Indianapolis, IN 46202-5132, USA
FEATURES
    Location/Qualifiers
    source
        1..1008
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
        223..816
        /note="similar to RhoLPs; geranylgeranylated protein"
        /codon_start=1
        /product="ATGP2"
        /protein_id="AAD00113.1"
        /db_xref="GI:4097563"
        /translation="MSASRFKICVTVGDAAGQYRRLKCTLLYSITNTPTDPTVDFNFG
        ANVYVGAATVNYGLDLDAGQEDYRRLRSTGEADVFLASLSISASTENYSGKNI
        ELKATPGPVIVLWCTKLLDRLDQFFIDHPGAVPPTAQGEALRLIGAPVYICS
        RTQENYGVFDAAIRVVLPPQKKKKSKAKQACSL"
CDS
    276 A 191 c 221 g 320 t
BASE COUNT
    ORIGIN

```

|    |     |                                                                |     |
|----|-----|----------------------------------------------------------------|-----|
| Qy | 2   | aaaaaacatgacgactcgcgaagattatcaacagtygtgcacgctcgggtgagcagctcggg | 61  |
| Db | 216 | AAGCAAAATGAGGCTCTCAGGTTCTAAAGGTGTGTCAKCTGTGGCAGCAGGCTGTGG      | 275 |
| Qy | 62  | gaaaattctgtatgctatctcatataacagcaacttctccaaaggtattgttccaac      | 121 |
| Db | 276 | TAAACCTGTTTGCTGATTCTTACACCGACACACTTTTCTACGGATTATGTACCGAC       | 335 |
| Qy | 122 | agttatttgataacttttagtccaaatgtggtgtgtagtcgacacagctgaaccttgcct   | 181 |
| Db | 336 | TGTTTTCGATACMTTATGCGCAAATGTGGTGTGATTGAAGCCACTGTGAATCTGGGCT     | 395 |
| Qy | 182 | atgggcactctgcgggcagaagaattataatagctaaaggccagcttaatatagaggagc   | 241 |
| Db | 396 | ATGGGATACCCGAGGACGAGAGGATTATACAACTATTAGACTCTTGATGTACCGCGTGC    | 455 |
| Qy | 242 | tgatgtgtgtttgttggcctttctcttataagcaaggccagttatgaaacactctcaaa    | 301 |
| Db | 456 | TGATGTTTTCATCTACGACTCTCTCTCTATCACTAGGCTAGTATGAGAAAGTCTTCAA     | 515 |
| Qy | 302 | aaagtggaatcccagactgaacacattatgtctcaatgatcacagcttgtctgtttggaac  | 361 |
| Db | 516 | GAGTGGATCCCGAGCTGAAGCATTTAGGCCCTGGTGTCCCTATAGTCTGTGTGGAAC      | 575 |
| Qy | 362 | caaacatgatttgcgagatgacacagcagcttctctctatgatacactggagcaacccaat  | 421 |
| Db | 576 | CAACATGATCTTCGGGAGACAAACAGCTTCTTCATTGACCAACCCGGCGCTGTACCAAT    | 635 |
| Qy | 422 | atcacacatctcaggggagaagactaaagagaatgataggcagcttacttatatagaattg  | 481 |
| Db | 636 | TACTACTCTCAGGAGGAGAACTGAAGAACTAATTGAGCTCCCGCATCATCTAGAGT       | 695 |
| Qy | 482 | cagctccaaaaccccaagaatgtgaaggtgttttgcgtgctgcaataaagaatgtttt     | 541 |
| Db | 696 | CAGTTCAAARAAACAAGAGAGCTGAAGAGGATTATTGATCAGCGATCCAGGTGGTCT      | 755 |
| Qy | 542 | gaggcCacacaaacccaagagaagactgtccaaaggagacatgtgtcttctcttgaat     | 601 |
| Db | 756 | TCAACTCTCCAAAGCAGAGAGAAAGAGAAACAAAGCACAACAAAGCTGCTCCATTTGTA    | 815 |
| Qy | 602 | att 604                                                        |     |
| Db | 816 | ATT 818                                                        |     |

RESULT 11  
AT041295  
LOCUS AT041295 1191 bp mRNA PLN 28-OCT-1997  
DEFINITION Arabidopsis thaliana GTP binding protein (ARAC1) mRNA, complete cds.  
ACCESSION U41295  
VERSION U41295.1 GI:1292907  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 1191)  
AUTHORS Winge, P., Brembu, T. and Bones, A.M.  
TITLE Cloning and characterization of rac-like cDNAs from Arabidopsis thaliana  
JOURNAL Plant Mol. Biol. 35 (4), 483-495 (1997)  
MEDLINE 98009984  
REFERENCE 2 (bases 1 to 1191)  
AUTHORS Winge, P., Brembu, T. and Bones, A.

Qy 7 caatgagccgcgaacattatacaagttgtctacagtcgtggtgagagctgtggggaaaa 56

Db 61 CAGAGATGGCGTCAGGCTTTATAGTGTGTGACGTCGGAGATGTCGCCGTGGAAAA 120

Qy 67 ctgttatgtctatctcatatatacagcaataacttcccaagcgtattgttcccaacgatat 126

Db 121 TCTGCATGCTCATTCTTACATAGCAATACTTTTCCATCTGATTATGTGCGCAACTGT 180

Qy 127 ttgtataactctttagtgcgaatgtgggtggtgagtgagcagcgaactgttgcctatggc 186

Db 181 TCGACAACTCTCAGTGTAAATGTGGTGTGTGAGCAACACTGTCAACTTGGATGTGGG 240

Qy 187 acactgcgggcagcagagattataataggtaaggccactgagtattatagagagctgagt 246

Db 241 ATACTCTGTCGACAGAGACTACACAACTGTAGACCTTTAGATTACCGTGCTGCTGAT 300

Qy 247 tgttttgttggctctctctctataaagcaagcgcgattatgaaaactctataaaaagt 306

Db 301 TTTTCATTCTTCTTCTCTCTCTATTACAGAGCTAGCTATGAGAAATAGCCAGAGAT 360

Qy 307 ggcataccagagctaaagccttattgtctataatgtaccgctgtgctcttggaacaaaac 366

Db 361 GGATTCCTGAGCTCAGCAATTATGCTCTGGTGTTCCCACTATCTGTTGGGACAAAAC 420

Qy 367 tagatttggcagatgaacagcgttctctcattgataccctggagcaaaccaatatca 426

Db 421 TGCATCTTGAGATGACAAGCAATCTTTATATGATCACTCTGGTGCCTGCCAATTACTA 480

Qy 427 catctacgggagagacactaaagagatgtatgagcgagtcattataagaatgcagct 486

Db 481 CAACACAGGAGAGAGCACTGAAGAACTGATTGGATCTGCTGCTCTACATGTGAATGTAGT 540

Qy 487 ccaaaacccaacgaagtgtgaagctgttttcgtatgctgcgaataaaagtattgtgagc 546

Db 541 CAAGACACACAGCAAGCTGAAGCGAGCTTTGATCACGATATAAAGATGGTGCCTCAGC 600

Qy 547 caccaaaaccaaagagaagccttgcgaagagcagactgtgcttcttgaattatggc 606

Db 601 CACCAACAGACAGAGAAGAGAAAGAAAGATAGAACCGTTGGCGCTCTTGTGATAGAAA 660

Qy 607 atctattatcagtcgaaaaagcgttaacaaagctgttgagat 650

Db 81 AAAGACCTGTGATGCTCATTTCTATACCAAGCATAGCTTCTCTACGAGTATGTTCCAA 140  
 Qy 122 agtatttgataacttttagtgcgaattgtgtgtgtgtgcgacgagtaacctgtgacct 181  
 Db 141 AGTTTTCGACAACTTCAGCGAAATGTGTGTGTGTGACGGGAGTACCGTGAACCTTGCCCT 200  
 Qy 182 atggagacctgcggggcgaagagattataatagcgtaagccactgagtatagagagac 241  
 Db 201 GTGGGATACTGCCGTCAGGAGATTAATATAGCTTAGGCTTTGAGCTACGAGAGAC 260  
 Qy 242 tgaatgtgtttttgtgaccttttcttataaagcagcaggtatgaasactctacaa 301  
 Db 261 AGATGCTCTCTATTATGCAATTTCCCTTATAGCAAGCCAGTACGAGATATTACAA 320  
 Qy 302 aaagtggatccagagctagacataatgtctataatgtaacgtgtgtgtgtgtgagac 361  
 Db 321 AAGTGGCTTCGGAGCTGAACATATGCTGTGCACTGCCCATTTGTCTGTGTGTGAC 380  
 Qy 362 caaacagattgttgagatgacagcagcttctctattgataccctggagcaaccacat 421  
 Db 381 AAAATGATGATTGAGGATGACAGCACTTCTTGAGGATCATCAAGAGCAGCTTCTAT 440  
 Qy 422 atcacactctcaggagagaaactaaagagatgagagcagcttacttatatagagat 481  
 Db 441 AACCACTGCTCAGGGAGAGAAATTAAGGAAATGATTGAGCTGTTAGGATCTTAGAGT 500  
 Qy 482 cagctccaaacacccacagatgtgagggctgttttcgattgctgataaagtagcttt 541  
 Db 501 CAGCTCCAAACCCACAGATGTGAAGCACTGTTGATACAGCATAGGCTAGCTTT 560  
 Qy 542 gagggccacaaacacaaagagagaag 566  
 Db 561 GAGGCCACAAAGGCCAAAGAAAG 585

RESULT 4  
 LTRAC2  
 LOCUS LTRAC2 982 bp mRNA PLN 12-MAY-1997  
 DEFINITION L. japonicus mRNA for small GTP-binding protein, RAC2.  
 ACCESSION Z73962  
 VERSION Z73962.1 GI:1370200  
 KEYWORDS rac2 gene; small GTP-binding protein.  
 SOURCE Lotus japonicus.  
 ORGANISM Lotus japonicus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Fabales; Fabaceae; Papilionoideae; Lotus.  
 REFERENCE 1 (bases 1 to 982)  
 AUTHORS Borg, S., Brandstrup, B. and Poulsen, C.  
 TITLE Structural analysis of cDNAs encoding 33 different small GTP binding proteins from Lotus japonicus and expression of corresponding mRNAs in developing root nodules  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 982)  
 AUTHORS Poulsen, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAY-1996) C. Poulsen, University of Aarhus, Dept of Molecular and Structural Biology, Gustav Wieds Vej 10C, DK-8000 Aarhus C, DENMARK  
 REFERENCE 3 (bases 1 to 982)  
 AUTHORS Borg, S., Brandstrup, B., Jensen, T.J. and Poulsen, C.  
 TITLE Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from Lotus japonicus, and expression of corresponding mRNAs in developing root nodules  
 JOURNAL Plant J. 11 (2), 237-250 (1997)  
 MEDLINE 97231879  
 FEATURES  
 source Location/Qualifiers  
 1..982  
 /organism="Lotus japonicus"  
 /variety="Gifu B-129"  
 /db\_xref="taxon:34305"  
 /tissue\_type="root nodules"

/dev\_stage="21 dpi with Rhizobium loti WZP 2037"  
 /clone\_lib="lambda ZAPII (Stratagene)"  
 113..703  
 /gene="rac2"  
 113..703  
 /gene="rac2"  
 /function="GTP-binding protein"  
 /codon\_start=1  
 /product="RAC2"  
 /protein\_id="CA98190.1"  
 /db\_xref="GI:1370201"  
 /db\_xref="SWISS-PROT:Q40220"  
 /translation="MSTARFKICVTVGDAVAKTKMLISTSTWPTFDVTVTVEDNFS  
 ANVVDGTVNLGLWDTAGQEDINRLPLSTAGADVFLAFLSLSRASYENISKWIP  
 ELRHATPTFVILVGLWDLREDKQILIDHFGATPTTGAESLAKAIGAAVILCES  
 KTKQNKAVFDAIKIVLPFKPKAKKKTKRPFVFL"  
 BASE COUNT 290 a 173 c 211 g 308 t  
 ORIGIN

Query Match 43.64; Score 396.8; DB 7; Length 982;  
 Best Local Similarity 78.84; Pred. No. 1.4e-71;  
 Matches 473; Conservative 0; Mismatches 127; Indels 0; Gaps 0;  
 Qy 6 acaatgagcaactgcagattatcaagtgtgtcagcgtgtgtgtgagctgtgggaaa 65  
 Db 110 AAAATGAGCAAGCTAGATTCACTCAAGTGTGTATCTGTGAGATGAGACGTGGGAAG 169  
 Qy 66 acttgatgtcttcttataaccagcaatactttcccaacgagattatgttccaaacagta 125  
 Db 170 ACCTGTATGCTTACTCTCTACACCGAACACATTCACCAAGGATATGTGCTACTGTT 229  
 Qy 126 ttgtataacttttagtgcgaattgtgtgtgtgtgtgagcagcagtaacctgtgacctatg 185  
 Db 230 TTGTATACTCAAGTGCATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 289  
 Qy 186 gcaactgtcggcggcgaagattataatgactgaagcgaactgagttatagagagctgat 245  
 Db 290 GACACTGTGCGAGGAGGATTACAAATAGCTTAGCGCTTGTGAGCTACAGAGGACAGAT 349  
 Qy 246 gtgtttttgtgaccttttcttataaagcagcgaattgaaaacactctacaaagag 305  
 Db 350 GTGTTCTGTGCTGCTTTTCCCTCTTACAGAGACGACTATGAAATATCTCCAAAGAG 409  
 Qy 306 tggatccacagcagtaagcaattatgtctcaatgtacaggtgtgtgtgtgtgtgtgtgt 365  
 Db 410 TGGATTCCTCACTGAGACCATATGCCCCAAGTGTGCCAATGTCTTCTGTGGGAACCAA 469  
 Qy 366 ctgactgttgagatgacagcagcttctctattgataccctggagcaacacactatca 425  
 Db 470 CTGTATTGAGGAGAGAGGAGGATTAATGATGATCTCTGAGGCCACACTATTACT 529  
 Qy 426 acatctcaggagagaagaactaaagagatgagagcagcttacttatatagagtcagc 485  
 Db 530 ACTGCCAGGAGAGAGAGCTGAAGAGCAATTTGGTGTCTGTGTGTACCTGAATGCAGC 589  
 Qy 486 tccaaacccacagcagtgatgagcgtgttttgcgtgtgtgtgtgtgtgtgtgtgtgtgt 545  
 Db 590 TCAAGATCTCAACAGATGTGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 649  
 Qy 546 ccacccaaacacagcagcgaactgtgcaaaagagaaactgtgttcttcttgaattattg 605  
 Db 650 CCACCTAAACCAAGAGAAAGACGAAGAGCAGACCATGCTGTTCTCTTAAATGTATG 709

RESULT 5  
 ATU49971  
 LOCUS ATU49971 843 bp DNA PLN 19-NOV-1998  
 DEFINITION Arabidopsis thaliana GTP binding protein Rop1at (Rop1at) mRNA, complete cds.  
 ACCESSION U49971  
 VERSION U49971.1 GI:2558665  
 KEYWORDS

|    |       |      |      |    |           |                    |
|----|-------|------|------|----|-----------|--------------------|
| 25 | 334.4 | 36.7 | 1008 | 8  | NTU64924  | U64924 Nicotiana t |
| 26 | 333.4 | 36.6 | 1117 | 8  | AF031428  | Arabidops          |
| 27 | 333   | 36.6 | 1220 | 8  | AF051223  | Picea mar          |
| 28 | 332.6 | 36.5 | 917  | 8  | MSA251210 | AJ251210 Medicago  |
| 29 | 330.8 | 36.4 | 850  | 49 | AF218381  | Oryza sat          |
| 30 | 329.8 | 36.2 | 1059 | 7  | AB024996  | AB024996 Cicer ari |
| 31 | 326   | 35.8 | 756  | 8  | ATUC6490  | U64920 Arabidopsis |
| 32 | 325.8 | 35.8 | 1081 | 49 | AF233447  | AF233447 Physcomit |
| 33 | 320.4 | 35.2 | 1393 | 8  | AF126053  | AF126053 Zea mays  |
| 34 | 317.2 | 34.9 | 1058 | 8  | AF126055  | AF126055 Zea mays  |
| 35 | 313.8 | 34.5 | 869  | 8  | AF079486  | AF079486 Arabidops |
| 36 | 310.6 | 34.1 | 640  | 49 | AF165925  | AF165925 Gossypium |
| 37 | 308   | 33.8 | 771  | 8  | AF156896  | AF156896 Arabidops |
| 38 | 291.8 | 32.1 | 1067 | 7  | AB029510  | AB029510 Oryza sat |
| 39 | 290.8 | 32.0 | 1045 | 8  | AF126054  | AF126054 Zea mays  |
| 40 | 290   | 31.9 | 734  | 8  | AF079485  | AF079485 Arabidops |
| 41 | 286   | 31.4 | 956  | 7  | AB029508  | AB029508 Oryza sat |
| 42 | 285.8 | 31.4 | 867  | 49 | AF239751  | AF239751 Tradescan |
| 43 | 283.2 | 31.1 | 1090 | 8  | AF079484  | AF079484 Arabidops |
| 44 | 282.2 | 31.0 | 1127 | 8  | AF126052  | AF126052 Zea mays  |
| 45 | 277.6 | 30.5 | 1087 | 7  | AB029509  | AB029509 Oryza sat |

## ALIGNMENTS

RESULT 1  
 S79308  
 LOCUS S79308 913 bp mRNA PLN 30-NOV-1995  
 DEFINITION Rac13-21.8 kDa GTP-binding protein [Gossypium hirsutum=cotton plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 913 nt].  
 ACCESSION S79308  
 VERSION S79308.1 GI:1087110  
 KEYWORDS  
 SOURCE upland cotton boll fibers cv. Acala SJ-2.  
 ORGANISM Gossypium hirsutum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
 REFERENCE 1 (bases 1 to 913)  
 AUTHORS Delmer,D.P., Pear,J.R., Andrawis,A. and Stalker,D.M.  
 TITLE Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibers  
 JOURNAL Mol. Gen. Genet. 248 (1), 43-51 (1995)  
 MEDLINE 95379748  
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbon 170155] from the original journal article.  
 This sequence comes from Fig. 1A.  
 FEATURES  
 source Location/Qualifiers  
 1..913  
 /organism="Gossypium hirsutum"  
 /db\_xref="taxon:3635"  
 gene 12..602  
 /gene="Rac13"  
 /note="21.8 kDa GTP-binding protein"  
 12..602  
 /gene="Rac13"  
 /note="21.8 kDa GTP-binding protein; pea Rho1 protein homolog/mammalian rac protein homolog; ; This sequence comes from Fig. 1A"  
 /codon\_start=1  
 /protein\_id="AA035093.1"  
 /db\_xref="GI:1087111"  
 /translation="MSTARFIVCVTVGGVGVKTCMLISYTNPTDVPVDFVDFNS ANVVVDSVNLGLWDTAGQEDYNRLAPLSYRGADVFLLASIKASYENITKNWIP ELRHAYNPVVLVGTGLDRDQKQDLDHGPATPISQGEHLKKMIGAVTITCSSL KIQQNVKAVDAIKVALRPPKPKRPPCKRRTCAPL"  
 CDS  
 307 a 169 c 172 g 265 t  
 BASE COUNT  
 ORIGIN

| Matches | 910; | Conservative                                               | 0;  | Mismatches | 0; | Indels | 0; | Gaps | 0; |
|---------|------|------------------------------------------------------------|-----|------------|----|--------|----|------|----|
| Qy      | 1    | aaaaaacatgagcactgcaagattatcaagtgtgtcaggtcggtgtgagcgtgtg    | 60  |            |    |        |    |      |    |
| Db      | 4    | AAAAACATGACNCTGCAAGATTATCAAGTGTGTACGGTGGGTGAGCATGTGG       | 63  |            |    |        |    |      |    |
| Qy      | 61   | ggaaacattgtatgtctctattatataccagcaactcttcccaacgattatgtccaa  | 120 |            |    |        |    |      |    |
| Db      | 64   | GGAAACTGTGATGTCATTTCATATACCAAGCAATCTTCCACGGATATGTCCAA      | 123 |            |    |        |    |      |    |
| Qy      | 121  | cagattttgataacttttagtcccaattgtgtgtggaaggcagcagcgaactgtgcc  | 180 |            |    |        |    |      |    |
| Db      | 124  | CAGATTGATAACTTTAGTCCCATGTGTGTGTGTGATGACGACGACGCAACCTGGCC   | 183 |            |    |        |    |      |    |
| Qy      | 181  | tatggacactgcggcgcaagaattataatagcttaagccactgattatgagag      | 240 |            |    |        |    |      |    |
| Db      | 184  | TATGGACACTGCCGGCAGAGATTATATAGGCTAAGGCCACTGAGTTATAGAGAG     | 243 |            |    |        |    |      |    |
| Qy      | 241  | ctgattgtttttgtgtgcttttctcttaagaacgcagttatgaacaacttaca      | 300 |            |    |        |    |      |    |
| Db      | 244  | CTGATGTTGTTTTGTGGCTTTTCTCTTATAGCAAGCGAGTATGAACAACTCTACA    | 303 |            |    |        |    |      |    |
| Qy      | 301  | aaagtgtgacccagcgtgaagcattatgctcaatgtaccagttgtctgtttgaa     | 360 |            |    |        |    |      |    |
| Db      | 304  | AAAGTGTGATCCAGAGCTAAGCATATGCTCATATGACCTAGGCTGATGTTGTGGA    | 363 |            |    |        |    |      |    |
| Qy      | 361  | ccaaactagatttggcagtgacaagcagttctctatgataccctggagacaaccaa   | 420 |            |    |        |    |      |    |
| Db      | 364  | CCAACTAGATTGGCAGATGACAGGACTTCTCATGTGACCTGGAGACACACAA       | 423 |            |    |        |    |      |    |
| Qy      | 421  | tatcaactctcagggaagaactaagaagatgataggacagttacttataagaat     | 480 |            |    |        |    |      |    |
| Db      | 424  | TATCAACTCTCAGGAGAGCACTAAGAAGATGATAGGACAGCTTACTATATAGAT     | 483 |            |    |        |    |      |    |
| Qy      | 481  | gcagctccaaacccaacagatgtgaagcgtgttttctgtatgtcgaataaagtagctt | 540 |            |    |        |    |      |    |
| Db      | 484  | GCAGCTCCAAACCCAAGATGTGAAGCGTGTCTTGATGCTGCAATAAAGTAGCTT     | 543 |            |    |        |    |      |    |
| Qy      | 541  | tgagccccaacaaacaaagaagaagcctgcaaaagaagaacattgtcttcttcaa    | 600 |            |    |        |    |      |    |
| Db      | 544  | TGAGCCCAACAAACAAAGAAGCGCTGCAAAAGAGACATGTGCTTCTTCCTTGA      | 603 |            |    |        |    |      |    |
| Qy      | 601  | tattgtatctattacagtcaaaacagtttaacaaagcgtgtgcagataaacatgaa   | 660 |            |    |        |    |      |    |
| Db      | 604  | TATTGATCTATTACAGTCAAAACAGTTAAACAAAGCTGTGCAGATAAACACTGAA    | 663 |            |    |        |    |      |    |
| Qy      | 661  | ctgtgatattgtttttttgttttacctatgttcacgtgaaactgaagaactcttaa   | 720 |            |    |        |    |      |    |
| Db      | 664  | TCGTGATAGTTTGTGTTTGTGTTTACATATGTTCCAGTGAACTATGAGACATCTTAA  | 723 |            |    |        |    |      |    |
| Qy      | 721  | gaaacccaactatcatataccactgatcatgaatgatgatttcaatttctgcagta   | 780 |            |    |        |    |      |    |
| Db      | 724  | GAANCCCAACTATCATATCAACCCATGATCAATGATGATTTCAATTTCCGAGTA     | 783 |            |    |        |    |      |    |
| Qy      | 781  | taagtctttttaaactcttttttttactcatatttaaacgaattctatgataatgtt  | 840 |            |    |        |    |      |    |
| Db      | 784  | TAGTCTCTTTTAACTCTTTCTTTTACTTCATTTTAAAGCAATTCATGATGAATGT    | 843 |            |    |        |    |      |    |
| Qy      | 841  | cctcaacacatgcttactcaagtttaataataatccattcttttacttaaaa       | 900 |            |    |        |    |      |    |
| Db      | 844  | CCTCAACATGCTTACATGATTTTAAATTAATAATCCATCTTCTTATTTACTAAA     | 903 |            |    |        |    |      |    |
| Qy      | 901  | aaaaaaaaa 910                                              |     |            |    |        |    |      |    |
| Db      | 904  | AAAAAAAAA 913                                              |     |            |    |        |    |      |    |

RESULT 2  
 S79309  
 LOCUS S79309 840 bp mRNA PLN 30-NOV-1995  
 DEFINITION Rac9-21.5 kDa GTP-binding protein [Gossypium hirsutum=cotton plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 840 nt].  
 ACCESSION S79309

Query Match 100.0%; Score 910; DB 8; Length 913;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-176;

04426.06



-----

[illegible]

```

ACCESSION   B08337.1
VERSION     B08337.1 GI:2089458
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
             Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
             Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
             Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1198)
AUTHORS     Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
             Ecker,J.
TITLE       BAC End Sequences at ATGC
JOURNAL     Unpublished (1997)
COMMENT     Other_GSSs: T19F9-T7.1, T19F9-T7, T19F9-Sp6
             Contact: Ecker J.
             Arabidopsis Thaliana Genome Center
             University of Pennsylvania
             Dept. of Biology, University of Pennsylvania, Philadelphia, PA
             19104
             Tel: 215-898-9384
             Fax: 215-898-8780
             Email: jecker@atgenome.bio.upenn.edu
             Seq primer: Sp6
             Class: BAC ends
             High quality sequence start: 48
             High quality sequence stop: 540.

FEATURES             source
             Location/Qualifiers
                 l..1198
                 /organism="Arabidopsis thaliana"
                 /strain="Columbia"
                 /db_xref="taxon:3702"
                 /clone="11939"
                 /clone_lib="TAMU"
                 /sex="hermaphrodite"
                 /note="Vector: pBelBACII; Site.1: HindIII; Site.2:
                 HindIII; Produced by Rod Wing"

BASE COUNT   98 a    101 c  108 g    757 t    134 others
ORIGIN

```

Query Match 9.0%; Score 86.8; DB 120; Length 1198;  
Best Local Similarity 42.8%; Pred. No. 4.9e-11;  
Matches 217: Conservative 0: Mismatches 290: Indels 0: Gaps 0:

[illegible]

Qy 589 gaaaaactgatgatgaggtataggaacataaacatgagtctccaaagcatgaaaaagagag 648  
 Db 672 AAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAA 613

Qy 649 gagaagaacactgagaaaggcagtagta 675  
 Db 612 NNAAAAAAAAAAAAAAAAAAAAAAAAAANA 586

```

RESULT 11
AQ330286/c
LOCUS      AQ330286      870 bp      DNA              GSS              08-JAN-1999
DEFINITION nhrb0046J18r CGI Rice BAC Library Oryza sativa genomic clone
            nhrb0046J18r, genomic survey sequence.

```

ACCESSION AQ30286  
VERSION AQ30286.1 GI:4122136  
KEYWORDS GSS.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Bukariyota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
REFERENCE 1 (bases 1 to 870)  
AUTHORS Wing,R.A. and Dean,R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: GGAACAGCAGTACGACCATG  
Class: BAC ends  
High quality sequence start: 13  
High quality sequence stop: 104.

```

FEATURES             Location/Qualifiers
     source            1..870
                        /organism="Oryza sativa"
                        /strain="Japonica"
                        /cultivar="Nipponbare"
                        /db_xref="taxon:4530"
                        /clone="nhb00463J18"
                        /clone_lib="CGI Rice BAC Library"
                        /tissue_type="Leaf"
                        /lab_host="E. coli DH10B"
     note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colono screening."

```

BASE COUNT 40 a 52 c 31 g 674 t 73 others

Query Match 8.9%; Score 85.8; DB 101; Length 870;  
Best Local Similarity 44.6%; Pred. No. 8.1e-11;  
Matches 225; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

|     |     |                                                              |     |
|-----|-----|--------------------------------------------------------------|-----|
| Dsb | 322 | AAATATATAAAAAAAAAAAAAAAAAATTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   | 381 |
| Qy  | 488 | acgaagagtagtcagagaagaacccgctgtccccaacgtgggaagcctaaggggcagc   | 547 |
| Dsb | 382 | AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | 441 |
| Qy  | 548 | agaaacaataaagccgaatatccgaaaatcctgagtgcaaggaaaactagatgagata   | 607 |
| Dsb | 442 | AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | 501 |
| Qy  | 608 | aggacaataaacatgctgtcccaaacgatgaanaagaggagagaacacctgagaag     | 667 |
| Dsb | 502 | AAAAAAAAAAKAKAGKBCDXAABKATTAATGGKKGAABASCBAAACATTTTSHBA      | 561 |
| Qy  | 668 | gcatagtaccctgagtggggttaaatgcctgaatgcgcgaagtcctatgtttactcgtct | 727 |
| Dsb | 562 | ACTIDKHAATTAATTSATAAAWAATTCABGTGVSBTSTSTATBSBTWTTTTATTT      | 621 |
| Qy  | 728 | ggctgcgcgaataagccttaagcctatgacactggtgcagtgcgcataatcctcagta   | 787 |
| Dsb | 622 | CBACCTTTTTACTATCTSMITSCATCTCYASTAGVSSATGCKSBTBAATKVMCWA      | 681 |
| Qy  | 788 | atttcctgggaattgttaattatat-tgtataaaaaagatggtagtggaanaatgctg   | 846 |
| Dsb | 682 | WTGVCYTITTAATTAATTTWATTTBTWTAAAMTTTATGAUKTSTVGVTISATTTTT     | 741 |
| Qy  | 847 | gtgtgcattcatccatgagca-atgctgaaatctcttgcatacagagattctcgaatgs  | 905 |
| Dsb | 742 | ACTCSBTTPPYACANCWSTAGTTSATNAWNSCATTTTSTTTGATTSCKKPCASNA      | 801 |
| Qy  | 906 | tataagttatgtatatatgtttctcagtcagtcagaataatttgatgtgtctgataag   | 965 |
| Dsb | 802 | TTTTTWTWTTGAAAGGCGAYTTTTTTTWTWTTTATTTTTTTTTTTWTAKCAASB       | 861 |
| Qy  | 966 | tt 967                                                       |     |
| Dsb | 862 | TT 863                                                       |     |

RESULT 7  
A0782441  
LOCUS A0782441 693 bp DNA GSS 02-AUG-1999  
DEFINITION HS\_3174\_A2\_B03\_WM\_C17 Approved Human Genomic Sperm Library D  
Homo sapiens genomic clone Plate=3174 Col=6 Row=C  
survey sequence.  
ACCESSION A0782441  
VERSION A0782441.1 GI:5685401  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,X., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99390589  
COMMENT On Mar 23, 1999 this sequence version replaced gi:3324197.  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 3174 row: C column: 6  
Seq primer: M13 Reverse

```

Class: BAC ends
High quality sequence stop: 693.

FEATURES             Location/Qualifiers
source                1..693
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-3174 Col-6 Row-C"
                     /clone_id="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones In
                     E-coli DH10B"

```

Query Match 9.1%; Score 87.8; DB 114; Length 693;  
Best Local Similarity 44.5%; Pred. No. 2.6e-11;  
Matches 218; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

|    |     |                                                               |     |
|----|-----|---------------------------------------------------------------|-----|
| QY | 168 | aaatatcgaaagacgcgcgaagctgtcgatactcaaacgcgaataatcatcgaaaggtacc | 227 |
|    |     |                                                               |     |
| Db | 175 | AAAAAANNAANAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAAAAANNAANNNN      | 234 |
| QY | 228 | aaatcgcggaagccgtgaatgtacgaaggaagaaacaaacctgcgaacatcatga       | 287 |
|    |     |                                                               |     |
| Db | 235 | AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAANNAANNA    | 294 |
| QY | 288 | agagtaccacgcagcttcgcgcatacgaaagcgcgcgaagctacgtataaagaaaacccga | 347 |
|    |     |                                                               |     |
| Db | 295 | AAAAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA      | 354 |
| QY | 348 | tttccccaattgggaagacgctaagaagcagcagagaacacagatcgaaattccgaaat   | 407 |
|    |     |                                                               |     |
| Db | 355 | AANNANNNANNNANNNANNNNNAANAAAAAAAAANAAAAAAAAAAAAAAAAAAAAA      | 414 |
| QY | 408 | accgcgagtacaagcaacaacgaatcgagaataagaacaataagatgaagatgccgaca   | 467 |
|    |     |                                                               |     |
| Db | 415 | AAANNAANNAANAAAAAAAAAAAAAAAAAAAAAAAAANNAANAAAAAAAAANNAANNA    | 474 |
| QY | 468 | gtcacgcgaatcgaaagcgcgcgaagctacgtacgagaagaaacccgcatttcccaaat   | 527 |
|    |     |                                                               |     |
| Db | 475 | NAAAAAAAAAAAAAAAAAAAAAAAAANNAANAAAAAAAAAAAAAAAAANNAANNAANNA   | 534 |
| QY | 528 | ggaaagccctaaggggccgcagaacacataagccgaatatccgcgaataacctgtgcga   | 587 |
|    |     |                                                               |     |
| Db | 535 | AAAAAAAAAAAAAAAAAAAAAAAAANNAANNAANAAAAAAAAAAAAAAAAANNAANNA    | 594 |
| QY | 588 | ggaaacacatgctgagataggaacataaacatgattcccaagcatgcaaaaagaga      | 647 |
|    |     |                                                               |     |
| Db | 595 | ANNAANNAANAAAAAAAAAAAAAAAAAAAAAAAAANNAANAAAAAAAAAAAAAAAA      | 654 |
| QY | 648 | ggagagaa                                                      | 657 |
|    |     |                                                               |     |
| Db | 655 | AAAAAAAAAA                                                    | 664 |

```

RESULT      8
CNS0122R
LOCUS       CNS0122R      839 bp      DNA              GSS              26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
              BACW01020 of DrosBAC library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION   AL01037
VERSION     AL01037.1  GI:5612648
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 839)
AUTHORS     Genoscope.
TITLE       Direct Submission

```

Db 634 AAGGAGGKGTSTWMCNCTWYKDDGGGAAAGTBASTWYTHWCKMKECTCAGAA 575  
 Qy 151 ctgcacaattgcttcacaatacagaagaacgcagagagctctgaatacaaacagccaaa 210  
 Db 574 MKTSTTGGKMTWGGGGGAAAGGKTTCTGAGAAAAAAAKKAAAAAAAAAAAAAA 515  
 Qy 211 tatctgaagagtgaccacaacatgagagcctggaatgtacaagggagaaaacaaaa 270  
 Db 514 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 455  
 Qy 271 cccctcaaacatctgaagagtgaccagagtcacgcgcgaatgcagagcagcagagatc 330  
 Db 454 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 395  
 Qy 331 gataaagaaaacccgatttccccaaatgggaaagcctaaagcagcagcagcagc 390  
 Db 394 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 335  
 Qy 391 gtgcatactcgaasaatcccgatcaaggaacaacagatgagataagaacataaa 450  
 Db 334 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 275  
 Qy 451 gatgaagagtgccagagtgacacagatcagaagcagcagagtgacagagaagaaa 510  
 Db 274 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 215  
 Qy 511 cccgatttccccaaatgggaaagcctaaagggcagcagcagcagcagcagcagc 570  
 Db 214 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 155  
 Qy 571 aaataacctgagtgcaaggaanaactagtgaggataaggaacataacatgagttccc 630  
 Db 154 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 95  
 Qy 631 aagctgaaaaaagagagagagaaaacctgagaaa 666  
 Db 94 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 59

RESULT 4  
 BL2963/c  
 LOCUS BL2963 759 bp DNA GSS 14-MAY-1997  
 DEFINITION T23D1-77.1 TAMU Arabidopsis thaliana genomic clone T23D1,  
 genomic survey sequence.  
 ACCESSION BL2963  
 VERSION BL2963.1 GI:2094085  
 KEYWORDS GSS.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;  
 Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 759)  
 AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and  
 Ecker, J.  
 TITLE BAC End Sequences at ATGC  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 15, 1999 this sequence version replaced gi:4123328.  
 Other\_GSSs: T23D1-Sp6.1, T23D1-Sp6, T23D1-77  
 Contact: Ecker J.  
 Arabidopsis Thaliana Genome Center  
 University of Pennsylvania  
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
 19104  
 Tel: 215-898-9384  
 Fax: 215-898-8780  
 Email: jecker@atgenome.bio.upenn.edu  
 Seq primer: 77  
 Class: BAC ends  
 High quality sequence start: 88  
 High quality sequence stop: 127.  
 FEATURES  
 Location/Qualifiers

source 1. 759  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="T23D1"  
 /clone\_lib="TAMU"  
 /sex="hermaphrodite"  
 /note="Vector: BeloBACII; Site\_1: HindIII; Site\_2:  
 HindIII; Produced by Rod Wing"  
 BASE COUNT 29 a 40 c 44 g 42 t 219 others  
 ORIGIN  
 Query Match 9.48; Score 90.8; DB 120; Length 759;  
 Best Local Similarity 33.74; Pred. No. 4.9e-12;  
 Matches 209; Conservative 0; Mismatches 412; Indels 0; Gaps 0;  
 Qy 194 aatacaaacgacccaattatcatgaagatcccaaacatgagagcctggaatgtaca 253  
 Db 756 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 697  
 Qy 254 agagagaaaaacacacccctgcaacatcatgaagagtgaccagctcgcgcagatga 313  
 Db 696 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 637  
 Qy 314 agagagcagcagagtgatcagaagaaaacgatttccccaaatgggaaagcctaaag 373  
 Db 636 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 577  
 Qy 374 agcagcagaacacagcagtgatcagaataccgagtgacagcagcagcagcagtg 433  
 Db 576 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 517  
 Qy 434 agataagaacacataagatgaagagtgccagagtgacacagcagtgagagcagcag 493  
 Db 516 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 457  
 Qy 494 agtcagagaagaaaacccgatttccccaaatgggaaagcctaaagggcagcagaa 553  
 Db 456 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 397  
 Qy 554 ataaagcgcagatccgaataactgagtgcaagaaaactagatgaggaaggac 613  
 Db 396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 337  
 Qy 614 ataacaatgatttcccaagcagataagaagagagagagaacactgagaagcgtatg 673  
 Db 336 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 277  
 Qy 674 taccctggtgggttmaastgctgaatggcgcagtcacatgttactcagctggtctg 733  
 Db 276 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 217  
 Qy 734 agcactagccttaagcctatgacactgtgcatgtgcatcatcatcagatatttca 793  
 Db 216 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 157  
 Qy 794 tgggattgtatattatattg 814  
 Db 156 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136

RESULT 5  
 CNS008G2  
 LOCUS CNS008G2 1101 bp DNA GSS 03-JUN-1999  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:  
 BACR35012 of RPECI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL073472  
 VERSION AL073472.1 GI:4953252  
 KEYWORDS GSS.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster









Qy 527 gggaaagccctaaagccagcagaaacataaaagccgaatccgaataactcctgagca 586  
 Db 470 GGGAAAGCCTTAAGGGCACGGAACATAAAGCCGAATATCCGAAATTAACCTGAGTCA 529  
 Qy 587 aggaaaactagtgagagataagcacaataaaactaggttcccaagcatgaaaaaag 646  
 Db 530 AGGAAAACTAGATGAGGTAAAGGAATAAATCAAGTGTCCCAAGCATGAAAAAGAG 589  
 Qy 647 aggaagaaacatgagagataagcacaataaaactaggttcccaagcatgaaaaaag 706  
 Db 590 AGGAAAGAACTAGATGAGGTAAAGGAATAAATCAAGTGTCCCAAGCATGAAAAAGAG 649  
 Qy 707 aagtcactgtttactagctggctgagcactaaagccttaagccatagacactgtgc 766  
 Db 650 AAGTCCAGTTTACTAGCTGTGGCTGAGCATAAGCCTTAAGCCATAGACACTGGTGC 709  
 Qy 767 atgtgcctatctatcagatatttctggatattgtatattatatttataaaaaag 826  
 Db 710 ATGTGCCTATCATCATGAGTAATTTATGGGATATTGTAAATATATTGTAAATAAAG 769  
 Qy 827 atgtgtgtgggaatgtgtgtgtgcctatccatg-agcaatgtgtaactctttgca 885  
 Db 770 ATGTGATGTGGGAATGTGTGTGTGATTCATCATGATAGCAATGCTGAATCTTTGCA 829  
 Qy 886 tgcagatagattctgagattgttattttatttatctctgtttgtctgagaaat 945  
 Db 830 TGCTAGAGATCTGAATGGTATTATGTTATGTTATCTGTTTGTCTAGTAATTAAT 889  
 Qy 946 ttgtaattgttattgtatattgt 967  
 Db 890 TTTGAATGTGTATGTAATGT 911

## RESULT 7

T43360  
 ID T43360 standard; DNA; 3974 BP.  
 AC T43360;  
 DT 11-MAR-1997 (first entry)  
 DE Cotton FbLate2-82A gene and promoter.  
 KW FbLate; promoter; fibre; transgenic plant; cotton; ds.  
 OS Gossypium hirsutum var. Sea Island.  
 FH Key Location/Qualifiers  
 FT promoter 1..2315  
 FT /\*tag= a  
 FT /note= "the FbLate promoter located between  
 FT bases 1 and 2315 is preferred for use in  
 FT constructs of the invention"  
 FT cds 2315..3379  
 FT /\*tag= b  
 FT /product= unidentified protein  
 PN W09639021-1.  
 PD 12-DEC-1996.  
 PF 06-JUN-1996; U09449.  
 PR 06-JUN-1995; US-467504.  
 PA (MONS.) MONSANTO CO.  
 PI John ME;  
 DR WPI: 97-042726/04.  
 PT Plant fibre-specific, developmentally regulated FbLate promoter -  
 PT useful for producing transgenic plants, esp. cotton, with altered  
 PT fibre properties  
 PS Claim 4; Page 57-59; 79pp; English.  
 CC A 3974 bp region (T43360) of clone pSK1FbLate2-28A includes the  
 CC fibre-specific FbLate promoter that is active during late fibre  
 CC development, plus a coding sequence (FbLate-82A; for an unknown  
 CC protein. The clone was obt'd. from a fibre genomic library using a  
 CC cDNA clone (see also T43362) that corresponds to RNA prevalent in  
 CC late fibre development, and insertion of an isolated clone into  
 CC Bluescript SK+ vector. The FbLate promoter can be used for tissue-  
 CC and developmental-specific expression of fibre and non-fibre  
 CC proteins (e.g. polyhydroxybutyrate biosynthetic enzymes) in  
 CC transgenic plants, esp. to alter the fibre characteristics of  
 CC cotton.  
 SQ Sequence 3974 BP; 1523 A; 603 C; 597 G; 1251 T;

Query Match 53.18; Score 513.4; DB 1; Length 3974;  
 Best Local Similarity 85.94; Pred. No. 7.1e-120;  
 Matches 587; Conservative 0; Mismatches 81; Indels 15; Gaps 1;

Qy 8 ttgttttaacacagtgctataaatttctgcatctcttctcttcttccaaactttactca 67  
 Db 2304 TTCGGTTAACCAAGGCTCATACACTGTGACCTTCTCTCTTTTCCAACTTTACTCA 2363  
 Qy 68 ttactgtctctaatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 127  
 Db 2354 TTATGTCTCTCATATGATGTGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 2423  
 Qy 128 acacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 187  
 Db 2424 AGACACAAACCACTCATCGAGTTGCCAATATAGTTCACAAATACGAAAGCGGAG 2483  
 Qy 188 agtct-----gaatacaaacagcaaaatattctgagagataccacaaac 232  
 Db 2484 AGTCGAATATGAAAGCCGGATACAAACGCCAAGATGACGAGAGTACCAAAAC 2543  
 Qy 233 atgagacacacacacacacacacacacacacacacacacacacacacacacacacacac 292  
 Db 2544 ATGACAGCCTGAAATGACAGGAGGAAACCAAAACCTGCACAGCACATGAGAGT 2603  
 Qy 293 accacagatcagcgcagatcagagagacagcagagatcagtaagaacacacacacac 352  
 Db 2604 AGACAGATGTCACCAATTCGAGGAGCAGCAGAGATGACGAGAGGAAACCGAGTTC 2663  
 Qy 353 ccaaatgggaagacacacacacacacacacacacacacacacacacacacacacacac 412  
 Db 2664 CCAATTTGGAAGACCTAAAGAGCAGAGAAACAGAGATGCAATATCCGAAATATCCG 2723  
 Qy 413 agtacaagacacacacacacacacacacacacacacacacacacacacacacacacac 472  
 Db 2724 AGTACAAGGAAACACAGATGAGGTGAGGACATAAATACGAGAGTACCAAGATCA 2783  
 Qy 473 acgaatcagacacacacacacacacacacacacacacacacacacacacacacacac 532  
 Db 2784 GTGATTCGAGGAGCAGCAGAGATACGAGAAAGAAACCGAGTTCGCCAATTTGAA 2843  
 Qy 533 agcctaagggcagcagacacacacacacacacacacacacacacacacacacacacac 592  
 Db 2844 AGCCTAAGAGCAGCAGGAAACACAGATGCAATATCCGGAATACCGAGATACAGGAA 2903  
 Qy 593 aactagatgagagataaggaacataaaactaggttcccaagcatgaaaaaagagagaga 652  
 Db 2904 AGCAGATGAGGATGAGGAACATAAATCAGGAGTGCCCAAGTCAACGAAATCGAAG 2963  
 Qy 653 agaaactcaggaagagcagtagta 675  
 Db 2964 AGCAAGAGAGTACGAGAAAGAA 2986

## RESULT 8

T43362  
 ID T43362 standard; cDNA; 645 BP.  
 AC T43362;  
 DT 11-MAR-1997 (first entry)  
 DE Cotton FbLate 2-82A gene cDNA clone All (FbLate-2).  
 KW FbLate; promoter; fibre; transgenic plant; cotton; ds.  
 OS Gossypium hirsutum.  
 PN W09639021-1.  
 PD 12-DEC-1996.  
 PF 06-JUN-1996; U09449.  
 PR 06-JUN-1995; US-467504.  
 PA (MONS.) MONSANTO CO.  
 PI John ME;  
 DR WPI: 97-042726/04.  
 PT Plant fibre-specific, developmentally regulated FbLate promoter -  
 PT useful for producing transgenic plants, esp. cotton, with altered  
 PT fibre properties

RESULT 4  
 T62624  
 ID T62624 standard; cDNA to mRNA; 1283 BP.  
 AC T62624;  
 DT 14-MAY-1997 (first entry)  
 DE Cotton fibre specific cDNA clone CXFB15-E9  
 KW cotton; fibre-specific; strength; transgenic plant; anthesis;  
 OS developmentally regulated; E5; H6; antisense; sense; ss.  
 NS Gossypium hirsutum strain Coker 312.  
 PN US559718-A.  
 PD 28-JAN-1997.  
 PF 04-OCT-1988; 253243.  
 PR 04-OCT-1988; US-253243.  
 RI 21-NOV-1990; US-617239.  
 FR 18-OCT-1993; US-138814.  
 FR 20-SEP-1995; US-530797.  
 PA (CETU ) AGRACETUS.  
 PI Brill WJ, John ME, Umbeck PF;  
 DR WPI; 97-108326/10.  
 PT Prodn. of transgenic cotton plants - by transformation with the H6  
 coding sequence or E5 anti-sense sequence, produces fibre of altered  
 strength  
 PS Example 4; Column 53-54; 33pp; English.  
 CC T62609-24 are cotton fibre-specific cDNA clones which can be used to  
 identify genomic clones. This clone, CXFB15-E9, is expressed in fibre  
 cells, but is also expressed at low levels in petal. (CX =  
 Coker; FB = Fibre; 10, 15 or 23 = age in days of fibre cells; A1 and the  
 last character and number stand for clone identity). The fibre-specific  
 genes were identified by differential cDNA library screenings. Coding  
 sequences from these isolated genes are used in sense or antisense  
 orientation to alter the fibre characteristics, e.g. strength, of  
 transgenic fibre-producing plants.  
 SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 75.5%; Score 730; DB 1; Length 1283;  
 Best Local Similarity 82.5%; Pred. No. 2.2e-174;  
 Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy 1 tttcttatttggtaaacatgggtcctataacttctgctacatttcttcttcttccaaact 60  
 Db 28 CTTTCTATTTGTAACCATGCGCTCATAACTTTTGTCATCTCTCTCTCTCTCTCTCT 87  
 Qy 61 ttactctattctgtctctctaatgtatgtatgacacacgtctgtctgacgtctgacat 120  
 Db 88 TTACTCATCTACTGTCTCTCACTAATAATCGGTAGTCACCGTCTGTGACGCGCTGACAT 147  
 Qy 121 ttattccacacacacacacactctcagatgctgccacattgtcttccaaatcagaaag 180  
 Db 148 TTATTCTAGACACAAACAACTCATCAGAGCTGCCAATTTGGCTTCAAAATCGAAAG 207  
 Qy 181 cagcagaggtctgaatacaaacagccaaattatcatgaaggtaccacacactgagag 240  
 Db 208 CACAAGAGTGTGTAATCAAAACCAACAAATATACGAAAGTACCAAAATCATGAGAG 267  
 Qy 241 cctgaatgtacagaggaggaacacacacacccctgcaacatcatgaaggtaccacag 300  
 Db 268 CTTAAATGTCACAGGAGGAGAAACAAACCCCTGCAACATCATGAAGGTACACAGAG 327  
 Qy 301 tctcgtcgtatgagagagcagcagaggtacgataaagacacccgatttcccccattg 360  
 Db 328 TCACCGGATTCGAGAGGACGACGAAGATACGATAAGAAAAACCCGATTCTCCCAATGG 387  
 Qy 361 gaaagcctaaagagcagcagagaaacacagattcgaatttcgaaataaccgagtcagag 420  
 Db 388 GAAAGGCTTAAGAGCACAGAAACACAGATTTGAATATCGGAAATACCGAGTACAG 447  
 Qy 421 gacaaacagagatg ..... 433  
 Db 448 GACAACAGAGTGGAGTAAGGAACATAAAATGAAGATACCATGAATACCGGATGCG 507

Qy 434 ..... 433  
 Db 508 AAGGAGCACAGAAATACGAGAAAGAAACCCGAGTTCGCCAAATGGGAAAGGCTAAA 567  
 Qy 434 ..... 433  
 Db 568 GAGCAGCAGAACCAAGATCGAATATCGGAAATACCCGAGTACAGGAAAGCAAGAT 627  
 Qy 434 -agataagaaacataaagtgaaggtgcagaggtgcagaggtgcagaggtgcagaggtgcagag 492  
 Db 628 AAGAGTAAGGACATAAAGTGAAGAGTGCCACAGGTGACACGATTCGAAAGTATCAGAA 687  
 Qy 493 gagtacgagaaagaaacccgatttcccaaatgggaagacctaaagggcagagaa 552  
 Db 688 GAGTACGAGAAAGAAACCAATTTCTCAATGGGAAAGCTTAAGAGCAGCGAGAA 747  
 Qy 553 cataaagcggatatacggaaatcctgagtgaaggaatactagatgaggtaaagaa 612  
 Db 748 CATAAAGCCGATATCAAAATATCCGAGTCAAGGAAAGCAAGTGAAGTAAAGAA 807  
 Qy 613 cataaacatgatttcccaagcagaaagaaagagagaaagcctgagaaagcagata 672  
 Db 808 GATAACATGAGTTCACAAAGCATGAAAGAGAGAGGAGAGAAACCTGAGAGAGCGAG 867  
 Qy 673 gtacccctgaggtgttaaatgctgatgacgaggtgacgttcttactgctgtgctc 732  
 Db 868 GTACCTCTGAGGTGTTAAATGCTGATGCGAGGATGATGTTTACTCAGCTGCGCTC 927  
 Qy 733 gagcactaagccttaagcactatgacactgtgcatgtgcatcatcatcagtcagtaatttc 792  
 Db 928 GAGCATATAGCTTAAGCATATGACATGCTGCTGATGCTCATCATCATGAGTAATTC 987  
 Qy 793 atgggtatgttatattattgttatataaaagatgtgaggaatgtgtgtgtgc 852  
 Db 988 ATGGGATATCGTAATTAATTTGTAATAAAGAGGTGGTGAATGGGAAATGTGTGTGTC 1047  
 Qy 853 attcatcatg-agcaatgctgaatctcttctgcatgcatgagattgtaagtgtatag 911  
 Db 1048 ATTATCCATGATGATGATCTGAATCTTTTGTGATGATGATGATGATGATGATGATG 1107  
 Qy 912 tttatgttatattgttctgtatgtaaatattttgaatgtgtatgtatgtt 967  
 Db 1108 TTTATGTAATATGTTGTTGCTAGTGAATTAATTTGAATGTGTATCATTAATTT 1163

RESULT 5  
 T70055  
 ID T70055 standard; cDNA; 1283 BP.  
 AC T70055;  
 DT 20-AUG-1997 (first entry)  
 DE Cotton fibre specific cDNA clone E9.  
 KW cotton; E5; fibre; promoter; transgenic plant; truncated;  
 NS heterologous gene expression; ds.  
 OS Gossypium hirsutum strain Coker 312.  
 PN US562082-A.  
 PD 15-APR-1997.  
 PF 04-OCT-1988; 253243.  
 PR 04-OCT-1988; US-253243.  
 RI 21-NOV-1990; US-617239.  
 FR 18-MAY-1992; US-885970.  
 FR 19-OCT-1994; US-298829.  
 PA (CETU ) AGRACETUS INC.  
 PI John M;  
 DR WPI; 97-235185/21.  
 PT DNA constructs contg. truncated promoter sequence - for  
 PT fibre-specific gene expression in cotton plants  
 PS Example 3; Column 45-48; 48pp; English.  
 CC T70040-57 are cotton fibre-specific cDNA clones which can be used to  
 obtain genomic clones containing fibre-specific promoters. Claimed DNA  
 constructs comprise a truncated promoter sequence (from one of T70031-38)  
 CC that promotes preferential gene expression in plant fibre cells, a  
 protein coding sequence not naturally associated with the promoter  
 CC sequence and a 3' termination sequence. The DNA constructs are useful for

Db 121 TATTCCTCACACACAAACACCTCTCAGAGCTGCACAAATGGCTTCAAAATACGAAAG 180  
 Oy 181 cagcagagctgtgaatacaacaacgcaaaatatctgaagagatcccaaaacatgagaag 240  
 Db 181 CACGAGAGCTGTGAATACAAACGCAAAATATCTATGAAGAGTACCCAAAATGAGAG 240  
 Oy 241 cctgaagattacagcagcagcaaaacaaacccctgcaaacatcatgaagagatccacag 300  
 Db 241 CCTGAATGTATCAAGGAGGAGAAACAAAACCCCTGCAACATCATGAAGAGTACCCAG 300  
 Oy 301 tcccgagctgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360  
 Db 301 TCACCGGAATCGAGGAGCAGCAGAGAGTACGATGAAGAAAACCCGATTTCCCAAAAG 360  
 Oy 361 gaaagcctaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420  
 Db 361 GAAAGCCTAAAGCAGCAGAGAAACAGAGTGAATATCCGAATATCCGAGTACAG 420  
 Oy 421 gcaaaacagctgtgaataaagaacaaagcagcagcagcagcagcagcagcagcagcagc 480  
 Db 421 GCAAAACAGTGTGAATTAAGAACATAAGATGAAGAGTACGAGGAGTACACAGATC 480  
 Oy 481 aaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540  
 Db 481 AAGGCGCAGAGAGTACGAGAGAAAACCCGATTTCCCAAAAGGAAAAGCCTAA 540  
 Oy 541 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 600  
 Db 541 GGGCAGCAGAAACATAAGACCGAATATCCGAAATACCTGAGTGCAGGAAAACATAG 600  
 Oy 601 gagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 660  
 Db 601 GAGGATAGAGAACTAAACATGATTTCCCAAGCATGAAAAGAGAGAGAGAAACCT 660  
 Oy 661 gagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720  
 Db 661 GAGAAGCATAGTACCTGAGTGGTTAAATGCTGAGTGGCAGTGCATGCTTAT 720  
 Oy 721 tgcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780  
 Db 721 TCAGCTGTGCTGAGCAGCAAGCCTTAAGCATATGACATGTGTCATGTGCATCATCA 780  
 Oy 781 tgcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840  
 Db 781 TGCAGTAAITTCATGGGATATTGTAATTATGTGTAATAAAAAAGTGTGAGTGGAA 840  
 Oy 841 atgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900  
 Db 841 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
 Oy 901 aatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960  
 Db 901 AATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960  
 Oy 961 taatgtt 967  
 Db 961 TAATGTT 967

## RESULT 2

T13048

ID T13048 standard; cDNA; 1283 BP.

AC T13048;

DT 27-MAY-1996 (first entry)

DE Cotton fibre-specific cDNA clone E9.

KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.

OS Gossypium hirsutum strain Coker 312.

PN US5495070-A

PD 27-FEB-1996.

PF 04-OCT-1988; 253243.

PR 04-OCT-1988; US-253243.

PR 21-NOV-1990; US-617239.

PR 18-MAY-1992; US-885970.

PA (CERTU ) AGRACERTUS INC.

PI John M.

DR WPI; 96-139095/14.

PT New isolated fibre-specific promoters - used for introducing

PS altered fibre-specific characteristics into plants, partic. cotton.

PT Example 3; Column 45-46; 48pp; English.

CC Cotton cDNA clone E9 (T13048) was isolated from a cDNA library of

CC cotton var. Coker 312 15-day-old boll cells using a subtractive

CC hybridization procedure. The clone hybridises strongly to fiber

CC RNA and weakly to petal RA. E9 and other fibre-specific cDNA clones

CC (see T13033-47 and T13049-T13050) were used to screen cotton genomic

CC libraries, leading to the isolation of genomic clones (see T13025-32

CC and T13052-53) contg. sequences capable of promoting gene expression

CC in fibre cells.

SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 75.5%; Score 730; DB 1; Length 1283;

Best Local Similarity 82.5%; Pred. No. 2.2e-174;

Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Oy 1 cttctctattgtttgaacccgttcataactttgtctctctctctctctctctctctctct 60  
 Db 28 CTTCTCTATTGTTGAACCATGGCTCATACTTTGTCATCTCTCTCTCTCTCTCTCTCTCT 87  
 Oy 61 ttactctattgtctctctctctctctctctctctctctctctctctctctctctctct 120  
 Db 88 TTACTCTATTGTTGCTCATTAATGCTGATGTCACACCGCTCTGTCAGCGCTGACAT 147  
 Oy 121 ttattccacacacacacacacacacacacacacacacacacacacacacacacacacac 180  
 Db 148 TTATTCGACGACAAACACCTCATCAGAGCTGCACAAATGGCTTCAAAATGAGAAAG 207  
 Oy 181 cagcagagctgtgaatacaacaacgcaaaatatctgaagagatcccaaaacatgagaag 240  
 Db 208 CACGAGAGCTGTGAATACAAACGCAAAATATCGAGAAAGTACCCAAAATGAGAG 267  
 Oy 241 cctgaagattacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300  
 Db 268 CTTAAATGACAGAGGAGAAAACAAAACCTGCAACATCATGAAGAGTACCAAGAG 327  
 Oy 301 tcccgagctgtgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360  
 Db 328 TCACCGGAATCGAGGAGCAGCAGAGAGTACGATGAAGAAAACCCGATTTCCCAAAAG 387  
 Oy 361 gaaagcctaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420  
 Db 388 GAAAGCCTAAAGCAGCAGAGAAACAGAGTGAATATCGAAAATACCGAGTACAG 447  
 Oy 421 gcaaaacagctgtgaataaagaacaaagcagcagcagcagcagcagcagcagcagcagc 480  
 Db 448 GACAAAAGATGAGGATGAAGAACATAAAATGAGAGTACCATGAATCAGCGAATCG 507  
 Oy 434 ----- 433  
 Db 508 AAGGAGCAGCAAGAAATACGAGAAAGAAAACCCGAGTTCGCCAAATGGGAAAAGCCTAA 567  
 Oy 434 ----- 433  
 Db 568 GAGCAGGAGAAACAGAGTGAATATCCGAAAATACCGAGTACAGGAAAAGCAAGAT 627  
 Oy 434 -agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 492  
 Db 628 AAGAGTGAAGAACATAAAGTAGAGAGTGCACAGAGTACACAGAAATCGAAGATCAAGAA 687  
 Oy 493 gactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 552  
 Db 688 GAGTCAGGAGAAAGAAAACCAATTTCTCAATGGGAAAAGCTTAAGGACAGCAGAA 747  
 Oy 553 catcaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 612  
 Db 748 CATTAAGCCGAATATCCAAAATACCGAGTGCAGAAAAGAAACAGATGAGGATGAAGAA 807

Db 3514 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3455

Qy 288 agagtaccagctgactgcgcgaatcgaggagcacyagaagtacgtataagaaancccca 347

Db 3454 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNNAAAAAAAAAAAAAA 3395

Qy 348 tttcccacaattgggaagcctaataagcgcagagaaacagcagtgctgaatatcgcaaat 407

Db 3394 AAAAAAAAAAAAAAAAAANNNAANNNAAAAAAAAAAAAAAAAANNNNANAANNAAAAAAAA 3335

Qy 408 acccgcttaccagtcacaaacagatgatgaataagaaatcataagatgaagtgctgcgga 467

Db 3334 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAA 3275

Qy 468 gtccacgcgaatcgaaagcagcagagtcgcgagaagaanaaacccgatctccccaaatg 527

Db 3274 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNNNNNAANNAAN 3215

Qy 528 ggaagtagctaaaggccagcagaacataagcgcgaatccgaatatcactgtgcga 587

Db 3214 AAAAAAAAAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNAANNAAAAAAA 3155

Qy 588 ggaanaacctagctgaggttaaggaacataacatgatgtcccaagcatgaanaagaga 647

Db 3154 AAAAAAAAAANNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3095

Qy 648 ggagagaacatctgagaaggcatgatctccctgatgggtttaaatgccctgaatgcgcg 707

Db 3094 AAAAAAAAAAACAAAAANNNNAANNNANNNANNNNNNNAGNNNNATTTATACCC 3035

Qy 708 agtccatctttactcagctctgcctgcagcactaagcctaagccatat 755

Db 3034 CCCCTGGTATGATGCATCTCTCAACCAAAAAAGATTATAAAT 2987

Search completed: September 3, 2000, 02:53:40  
Job time: 31570 sec

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |            |                 |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----------------|
| RESULT 15  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |            |                 |
| AC013349.C |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |            |                 |
| LOCUS      | AC013349 129404 bp                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | DNA        | HTG 06-FEB-2000 |
| DEFINITION | Homo sapiens clone RP11-22K1, LOW-PASS SEQUENCE SAMPLING.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |            |                 |
| ACCESSION  | AC013349                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |            |                 |
| VERSION    | AC013349.2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | GI:6910730 |                 |
| KEYWORDS   | HTG, PHASD0.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |            |                 |
| SOURCE     | human.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |            |                 |
| ORGANISM   | Homo sapiens                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |            |                 |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |            |                 |
| REFERENCE  | 1 (bases 1 to 129404)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |            |                 |
| AUTHORS    | Birren,B., Linton,L., Nusbaum,C. and Lander,E.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |            |                 |
| TITLE      | Homo sapiens, clone RP11-22K1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |            |                 |
| JOURNAL    | Unpublished                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |            |                 |
| REFERENCE  | 2 (bases 1 to 129404)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |            |                 |
| AUTHORS    | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckley,R., Boguslavsky,L., Bouckhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardysa,S., Grant,G., Hagos,B., Haeflora,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karasata,A., Klein,J., Lechoksky,J., Liew,C., Locke,K., Macdonald,P., Margus,N., McEwan,P., McGurk,A., McKernan,C., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,A., Tesfaye,S., Tirrell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W., Zimmer,A. and Zody,M. |            |                 |

TITLE Direct Submission  
JOURNAL Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Feb 6, 2000 this sequence version replaced gi:6272406.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seg.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

```
----- Project Information
Center project name: L4134
Center clone name: 22_K_1
```

- \* NOTE: this record contains 151 individual
- \* sequencing reads that have not been assembled into
- \* contigs. Runs of N are used to separate the reads
- \* and the order in which they appear is completely
- \* arbitrary. Low-pass sequence sampling is useful for
- \* identifying clones that may be gene-rich and allows
- \* overlap relationships among clones to be deduced.
- \* However, it should not be assumed that this clone
- \* will be sequenced to completion. In the event that
- \* the record is updated, the accession number will
- \* be preserved.

|   |      |       |                            |
|---|------|-------|----------------------------|
| * | 1    | 930:  | contig of 930 bp in length |
| * |      |       | gap of unknown length      |
| * | 931  | 1855: | contig of 925 bp in length |
| * |      |       | gap of unknown length      |
| * | 1856 | 2759: | contig of 904 bp in length |
| * |      |       | gap of unknown length      |
| * | 2760 | 3706: | contig of 947 bp in length |
| * |      |       | gap of unknown length      |
| * | 3707 | 4566: | contig of 860 bp in length |



```
repeat_region 22788..22924
                /rpt_family="B3"
repeat_region 22938..23864
                /rpt_family="RLRUI3C"
repeat_region complement(23896..23934)
                /rpt_family="(GAAA)n"
repeat_region 25028..26295
                /rpt_family="MMETN"
repeat_region 26296..26443
                /rpt_family="B1_MM"
repeat_region 26445..26523
                /rpt_family="POLY_A"
repeat_region complement(26524..26782)
                /rpt_family="RLRUI"
repeat_region complement(26827..26953)
                /rpt_family="RLRUI"
```

```

Query Match          9.2%; Score 89.4; DB 12; Length 149886;
Best Local Similarity 50.1%; Pred. No. 4.3e-09;
Matches 249; Conservative 0; Mismatches 246; Indels 2; Gaps 1;

Qy 173 acgaagaacagcagagagtctgtgatacaaaacagccaaaatctcatgaaggtaccccaac 232
Db 12625 AGGAATAGAGAGAGAGAGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 12684

Qy 233 atgagagcctgaatatgtacaaggaggagaaaacaaacctgcacaaatcatgaagagt 292
Db 12685 AGGAAAAAAGAGAGGAGGAGGAGGAGGAAGAGAGAGAGAGAGAGATTTGAGAGAGGAG 12744

Qy 293 accacagctgcagtcggaatgcgaagagcagcaagagtcagtaagaaaaacccgatttcc 352
Db 12745 AGGAGAGAGAGAAAGAGGTGGAGGATGAGGAGAGAGAGAGAGAGAGAGGAGGAGAGAGAG 12804

Qy 353 ccaaatgtggaaaamgctaaagagcagcgagaaacacgaagtcgaatatccgaagaatcccg 412
Db 12805 AGGAGAGAGAGAGAGAGAGAGAGGAGGAGAGATAGATGAGGAGAGAGAGAGAGAGAGAG 12864

Qy 413 agtacaagagcaaacagagtagagataagaacaataaagtagaagagtgccgaagtgatcc 472
Db 12865 AGGAGAGAGAGAGAGAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 12924

Qy 473 acgaagct--aaagagcgcgaagagtcagagaaagaaaacccgatttcccccaattggga 530
Db 12925 AGGAGAGAGAGAGAGAGAGAGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 12984

Qy 531 aaagcctaaaggggccagagaacataaagccgaatatctcgaatatccttgatgtcagga 590
Db 12985 AGGAG 13044

Qy 591 aaagcagtagatgatgagagagcaacataacgtgtctccaaagagtgaaagaaagagga 650
Db 13045 CAAATGTGGAGGAGAGAGAGAGAGAACTTTGAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAG 13104

Qy 651 gaagaaacctgagaaag 667
Db 13105 GAGGAGAGAGAGAGAGAG 13121

```

```

RESULT 11
CEL3368
LOCUS      CEL3368      39103 bp      DNA              INV      08-JUL-1998
DEFINITION Caenorhabditis elegans cosmid C3368.
ACCESSION  U53154
VERSION     U53154.1 GI:1255414
KEYWORDS
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditidae; Rhabditidae; Peldicerinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 39103)
AUTHORS     Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,

```

```

/db_xref="taxon:9606"
/clone="84_E_24"
/clone_lib="Alan Buckler -- per comm"
/map="17"
/chromosome="17"
repeat_region 446..753
/rpt_family="AluSc"
repeat_region 2375..2419
/rpt_family="(TTTTC)n"
complement(2985..3103)
/rpt_family="L2"
repeat_region complement(3146..3352)
/rpt_family="L2"
repeat_region complement(3531..3639)
/rpt_family="L2"
repeat_region complement(3798..4394)
/rpt_family="L2"
repeat_region 5531..5631
/rpt_family="MER81"
repeat_region 5648..5674
/rpt_family="(CAAAA)n"
repeat_region 5812..6020
/rpt_family="L1MC4"
repeat_region 6180..6224
/rpt_family="AT-rich"
repeat_region 6303..6611
/rpt_family="AluSp"
repeat_region 7378..7618
/rpt_family="MIR"
repeat_region 9593..9623
/rpt_family="(GGGA)n"
complement(10666..10923)
/rpt_family="AluSq"
repeat_region 10958..11088
/rpt_family="MIR"
repeat_region 11752..11914
/rpt_family="L2"
complement(11915..12219)
/rpt_family="AluSx"
repeat_region 12220..12347
/rpt_family="L2"
repeat_region 13110..13182
/rpt_family="GA-rich"
repeat_region 13526..13564
/rpt_family="(TCC)n"
repeat_region 13575..13660
/rpt_family="(TA)n"
complement(14380..14658)
/rpt_family="AluJo"
repeat_region 14742..14830
/rpt_family="(TA)n"
repeat_region 14860..14967
/rpt_family="(TA)n"
complement(14977..15122)
/rpt_family="L2"
repeat_region complement(15292..15385)
/rpt_family="L2"
repeat_region 16672..16781
/rpt_family="L2"
repeat_region 16802..16846
/rpt_family="(TGA)n"
repeat_region 17225..17245
/rpt_family="(A)n"
repeat_region 18411..18720
/rpt_family="AluJb"
complement(20715..20788)
/rpt_family="MLT137"
repeat_region complement(20822..20916)
/rpt_family="MLT137"
repeat_region 20956..21137
/rpt_family="MIR"
repeat_region 21830..21872
/rpt_family="AT-rich"

```

```

repeat_region 21957..22257
/rpt_family="AluSx"
repeat_region complement(22438..22538)
/rpt_family="MLT1"
repeat_region complement(23375..23667)
/rpt_family="AluY"
repeat_region 24553..24858
/rpt_family="AluY"
repeat_region complement(25021..25234)
/rpt_family="MIR"
repeat_region 26557..27261
/rpt_family="AluY"
repeat_region complement(27524..27669)
/rpt_family="MLT1B"
repeat_region complement(27670..27964)
/rpt_family="AluSp"
repeat_region complement(27965..28210)
/rpt_family="MLT1B"
repeat_region complement(28228..28522)
/rpt_family="AluSq"
repeat_region complement(28592..28700)
/rpt_family="MIR"
repeat_region complement(29353..29537)
/rpt_family="MIR"
repeat_region 29802..30095
/rpt_family="AluSq"
complement(30099..30219)
/rpt_family="MIR"
repeat_region 30227..30267
/rpt_family="(TG)n"
repeat_region 30276..30321
/rpt_family="(CA)n"
complement(30322..30511)
/rpt_family="AluJo"
complement(30956..31151)
/rpt_family="MER3"
complement(31533..31633)
/rpt_family="L2"
repeat_region 31810..31858
/rpt_family="L2"
repeat_region 32028..32107
/rpt_family="L1PA4"
complement(32919..33182)
/rpt_family="MIR"
repeat_region 34169..34314
/rpt_family="CT-rich"
repeat_region 34774..35007
/rpt_family="MIR"
repeat_region 35589..35621
/rpt_family="AT-rich"
repeat_region 35763..36142
/rpt_family="MER41A"
repeat_region 36143..36452
/rpt_family="AluSq"
repeat_region 36453..36636
/rpt_family="MER41A"
repeat_region 38011..38071
/rpt_family="CT-rich"
repeat_region 38083..38262
/rpt_family="(TTC)n"
repeat_region 38235..38319
/rpt_family="(TTC)n"
repeat_region 38320..38499
/rpt_family="(TTC)n"

```

```

Query Match          9.4%; Score 91; DB 39; Length 180385;
Best Local Similarity 48.9%; Pred. No. 2e-09;
Matches 244; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

Qy 169 aaatagaaaagacgaagagtcgaatacaaacgccaataatcatgaagatcaccc 228
   || ||||| ||||| || || || || || || || || || || || || || || ||
Db 38579 AAGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 38520

```



Ov. 263. aacaaaaaccctgcaaacatcatgaagagtaccacgaatcacccgaatcgaaggagcacc 322

```

RESULTS 8
AB001373
LOCUS      AE001373      12029 bp      DNA      INV      06-NOV-1998
DEFINITION Plasmidium falciparum chromosome 2, section 10 of 73 of the complete sequence.
ACCESSION AE001373 AE001362
VERSION   AE001373.1 GI:3845097
KEYWORDS
SOURCE     malaria parasite P. falciparum.
ORGANISM   Plasmodium falciparum
Bukarya: Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 12029)
AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravid,L., Rooni,Z.V., Shailon,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lal,Z., Schwartz,D.C., Pertea,K., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
TITLE      Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL   Science 282 (5391), 1126-1132 (1998)
MEDLINE   99021743
REMARK     Erratum:[[published erratum appears in Science 1998 Dec 4;282(5395):1827]]
REFERENCE  2 (bases 1 to 12029)
AUTHORS   Gardner,M.J.
TITLE      Direct Submission
JOURNAL   Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA
FEATURES   Location/Qualifiers
source     1. 12029
            /organism="Plasmodium falciparum"
            /db_xref="taxon:5833"
            /chromosome="2"
gene       3709..5236
            /gene="PF8B0110w"
            join(3709..4318,5079..5236)
            /gene="PF8B0110w"
            /note="predicted by GlimmerW"
            /codon_start=1
            /product="predicted integral membrane protein"
            /protein_id="AAC71812.1"

```

Qy 61 ttactcattactgtctcactaatgatcggtagccacacggtctcgtcagggctcgacat 120  
Db 88 TTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCGTCAGGGCTCGACAT 147

```

RESULT 5
GBU34401
LOCUS GBU34401 1699 bp DNA PLN 01-JAN-1996
DEFINITION Gossypium barbadense FbLate-2 gene, complete cds.

```

c 25 83.2 8.6 76897 42 AC016179 Homo sapi  
 26 83 8.6 169931 11 AC005822 Homo sapi  
 c 27 82 8.5 162575 41 AC004086 Homo sapi  
 c 28 81.6 8.4 63440 53 AC024285 Homo sapi  
 c 29 79.6 8.2 80432 51 AC022680 Homo sapi  
 c 30 79.6 8.2 222193 32 CNS01DS8 Homo sapi  
 31 79.4 8.2 34488 82 KSU52054 Homo sapi  
 c 32 79.4 8.2 32207 5 AR065852 Arabidops  
 c 33 79.4 8.2 137508 82 KSU75698 Arabidops  
 c 34 79.4 8.2 141753 49 AC009323 Arabidops  
 c 35 79.4 8.2 174383 41 AC009781 Arabidops  
 36 78.8 8.1 42839 69 ACU27282 Homo sapi  
 37 78.6 8.1 164520 43 ACU20738 Homo sapi  
 38 78.2 8.1 2000 2 AF019082 Homo sapi  
 39 78.2 8.1 27323 2 AB000789 Homo sapi  
 40 78.2 8.1 110000 31 PFMAL4P1\_0 Plasmodi  
 41 78.2 8.1 160547 41 HS1164110 Homo sapi  
 42 77.8 8.0 43907 34 CELF36H12 Homo sapi  
 c 43 77.8 8.0 119500 51 ACU15927 Homo sapi  
 c 44 77.6 8.0 89072 55 ACU25070 Homo sapi  
 c 45 77.2 8.0 73020 51 AC02851 Homo sapi

## ALIGNMENTS

RESULT 1  
 I18362  
 LOCUS I18362 1283 bp DNA PAT 07-OCT-1996  
 DEFINITION Sequence 17 from patent US 5495070.  
 ACCESSION I18362  
 VERSION I18362.1 GI:1598717  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCES 1 (bases 1 to 1283)  
 AUTHORS John.M.  
 TITLE Genetically engineering cotton plants for altered fiber  
 JOURNAL Patent: US 5495070-A 17 FEB-1996;  
 FEATURES Location/Qualifiers  
 source 1..1283  
 /organism="unknown"  
 BASE COUNT 509 a 233 c 251 g 290 t  
 ORIGIN

Query Match 75.5%; Score 730; DB 5; Length 1283;  
 Best Local Similarity 82.5%; Pred. No. 4e-140;  
 Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy 1 cttctctatttggttaaccattggtctataactttgtctatctcttctcttttccaaatt 60  
 Db 28 CTTCTCATTTGTGTAACCATGGCTCATACATTTTGTCACTCTTCTCTCTTCTCACTT 87  
 Qy 61 ttactctattctgtctctactaactgctggtgagccacgctgtctgctcggcgtgcacat 120  
 Db 88 TTACTCATTTACTCTCTCACTAATAATGGTACTCACACCGCTCTGTCACGGCGTGCAT 147  
 Qy 121 ttattccacacacaaacactctctcagagctgtcccaatttggtctcaaatcagaaag 180  
 Db 148 TTATTCGACGACAAACACACCTCATCAGAGCTGCCCAATTGGCTTCAAAATACGAAAG 207  
 Qy 181 cagcagagctctgaatcacaacagccaaatatctgagaggtaccacaacatcagagag 240  
 Db 208 CACAAAGAGTCTGAATACAAACACCAAAATATACGAAAGTACCCAAACATCAGAGAG 267  
 Qy 241 cctgaattgtagcaggagagaaacacaaacccctgaacacatctgagagtagcacagag 300  
 Db 268 CCTAAATGTCACAGGAGGAAACAAACAAACCTCTCAACATCATGAGAGTACCAAGAG 327  
 Qy 301 tcacgagatcgaggagcagcagagtagcagataagaagaaacccgatttcccaattgg 360  
 |||||

Db 328 TCACGCGAATGAGGAGGACGACGAGATGACGATTAAGAAAAACCGATTTCCTCCAAATGG 387  
 Qy 361 gaaagcctaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 420  
 |||||  
 Db 388 GAAAGCCTAAGAGCGCCAGAAACACGAGTTGATATATCCGAAATACCCGAGTACAG 447  
 Qy 421 gacacacagatg----- 433  
 |||||  
 Db 448 GACACACAGATGAGGATGAAGACATAAAAATGAGAGTACCATGAATCAGCGAATCG 507  
 Qy 434 ----- 433  
 Db 508 AAGGAGCAGCAAGAAATACGAGAAAGAAAAACCGAGTTCCCAATGGGAAAGCCTAAA 567  
 Qy 434 ----- 433  
 Db 568 GAGCACGAGAAACACGAGTGAATATCCGAAATACCCGAGTACAGGAAAGCAGAT 627  
 Qy 434 -agataagaacataaagatgaagagtcacagagtcacagcagcagcagcagcagcag 492  
 Db 628 AAGAGTAGGACATAAAGATGAAGAGTCCAGAGTCACAGCAATGAAAGTACGAG 687  
 Qy 493 gagtacgagagagaaacccgatttcccaattgggaaagcctaaggcgacagagaa 552  
 Db 688 GAGTACGAGAAAGAAAAACCAATTTCTTCAATGGGAAAGCCTAAGAGCAGCAGAA 747  
 Qy 553 cataagcgcagatccgaaatcactgagtcgaagagaaacatagatgaggaagaa 612  
 Db 748 CATAAAGCCGATATCCAAAAATACCGAGTCAAGGAAACACAGATGAGATAGAA 807  
 Qy 613 cataaacatgagtcccaagcagaaagagagagagaaacatgagaaagccta 672  
 Db 808 GATAAATCATGAGTCCCAAGCATGAAAGAGAGAGAGAGAAACCTGAGAAAGCGAGA 867  
 Qy 673 gtacccgtgaggtgttaaatgctcgaatgctgagagcagcagcagcagcagcagc 732  
 Db 868 GTACCTGATGGGTGTAATATGCTGAATGCGCAGATCCATGTTTCACTCATCTGCTC 927  
 Qy 733 gacactaagccttaagcagatgacactggtgagtcgcatcatcatcagagtaattc 792  
 Db 928 GAGCATTAGCTTTAAGCATATGACATGTCGTGATGTCATCATCATGCAATTTTC 987  
 Qy 793 atgggtattgttaattattgtttaaaagagtggtgagggaaatgtgtgtgtgc 852  
 Db 988 ATGGGATATCGTATATATATTTGTTAATAAAAGATGGTGAGTGGGAAATGTGTGTGC 1047  
 Qy 853 attctccatg-agcactgctgaatctcttctgctgcatgagattcgaatgttatag 911  
 Db 1048 ATTATCATCATATGACATGCTGAATCTCTTTCATGATGAGATTTCTGAAGATTATAG 1107  
 Qy 912 ttattgttatatgtttgtttgtatgaaattattttgaatgtgtatgattgt 967  
 Db 1108 TTTATGTTATATCGTTGTTCTAGTGAATTAATTTTGAATGTGTATCATTAATGT 1163

RESULT 2  
 I21349  
 LOCUS I21349 1283 bp DNA PAT 07-OCT-1996  
 DEFINITION Sequence 17 from patent US 5521078.  
 ACCESSION I21349  
 VERSION I21349.1 GI:1601703  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCES 1 (bases 1 to 1283)  
 AUTHORS John.M.  
 TITLE Genetically engineering cotton plants for altered fiber  
 JOURNAL Patent: US 5521078-A 17 28-MAY-1996;  
 FEATURES Location/Qualifiers  
 source 1..1283  
 /organism="unknown"  
 BASE COUNT 509 a 233 c 251 g 290 t



RESULT 9  
 US-08-931-999-4  
 Sequence 4, Application US/08931999  
 Patent No. 6043219  
 GENERAL INFORMATION:  
 APPLICANT: Iandolo, John J.  
 APPLICANT: Crupper, Scott S.  
 TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Hovey, Williams, Timmons & Collins  
 STREET: 2405 Grand Boulevard, Suite 400  
 CITY: Kansas City  
 STATE: Missouri  
 COUNTRY: U.S.A.  
 ZIP: 64108  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/931,999  
 FILING DATE:  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/710,561  
 FILING DATE: 19-SEP-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Collins, John M.  
 REGISTRATION NUMBER: 26,262  
 REFERENCE/DOCKET NUMBER: 25043-A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 816/474-9050  
 TELEFAX: 816/474-9057  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6755 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Staphylococcus aureus  
 STRAIN: UT0007  
 US-08-931-999-4

```

RESULT 10
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5765597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release 1.0, Version 1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991

```

Qy 263 aacaaataatgcgaactatgaagattaccacgagctacccgatacgaagagcaacg 322  
Db 1507 AAAAAGCGCATAGCGCATCTAGTATTTTACCTATCTATCAAGTATGTAAGGATAC 1448

Qy 323 aagatagcgtataaagaaaaccgcatttcccataaggaaaagcttaagagcgcagca 382  
Db 1447 AAGATTTTGGTCAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1368

Qy 383 aacacgaagtcgataatccgaataatcccgagtcacagcacaacagatgagataaga 442  
Db 1387 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1328

Qy 443 aacataaagctgaagagctccagagtagcacaagctgaagaagcagcagaagctacgaa 502  
Db 1327 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1268

Qy 503 aagaaaataccgcatttcccataatggaaaagcctaaggcgcagaagaataagacgc 562  
Db 1267 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1208

Qy 563 aatatccgaaataacctgagtcaggaagaaactagatgagcagaagcaataacatcg 622  
Db 1207 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1148

Qy 623 agttcccagaagctgaaaaggaagagagaagaaactgagaagcatagctacctgag 682  
Db 1147 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1088

Qy 683 tgggttaaaatgcctgataagtg 703  
Db 1087 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1057

RESULT 6  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHRIFLINGER, F.  
APPLICANT: PALMER, P. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Poley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria

RESULT 4

|    |     |                                                                |     |
|----|-----|----------------------------------------------------------------|-----|
| Qy | 1   | ctttctattgtgttaacacagctcaataacttctgtcaatcttcttctttccaaact      | 60  |
| Db | 28  | CTTTCATTCTTGTAACCACTGGCTCAAACTTTTGATCTCTTCTTCTCTTCCAACTT       | 67  |
| Qy | 61  | ttaactcaattactgtctcactatgatcgtgtagcacacccgtctcgtcagcggctgcacat | 120 |
| Db | 88  | TTACTCAATTACTGTTCTCAATAAATCGGTAGTGCACACGCTCTGCTCAGCGGCTGCACAT  | 147 |
| Qy | 121 | ttaattcacacacaaacactcactcaacagctctccacacttgctctcgaatacgaagaa   | 180 |

Db 868 GTACCTTGAGTGGGTAAATGCTTGAATGCCGAATCCATGTTTACTCAGTCTGGCTC 927

Qy 733 gagcactaagccttaagccatatgacactggatgtgccatcatcatgcagtaatttc 792

Query Match 75.5%: Score 730: DB 1: Length 1283:



00000000



Pterygota, Neoptera, Endopterygota, Diptera, Brachycera,

Db 576 TTTTTTTTTTTATKTATAAAAKTWWAAAAAAWTTTTTTTTTTWTTKTTTTTTWTTTAT 517

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Goezawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://dncac.med.buffalo.edu/drosophila/bac.htm>.

```

FEATURES             Location/Qualifiers
     source            1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR08K08"
                        /note="end : TET3"
BASE COUNT            395 a    120 c    103 g    334 t    149 others
ORIGIN

```

Query Match 3.9%; Score 118.4; DB 122; Length 1101;  
Best Local Similarity 42.2%; Pred. No. 1.5e-08;  
Matches 234; Conservative 61; Mismatches 260; Indels 0; Gaps 0;

Qy 383 aactttgataagtcaccaaactttaacaagaatttgattgtgatcatatatatatatct 442  
| |||| | || : | ||: : || | | | | : |  
pb 510 ACCTTAATATAAAGAAAAAAWAAAAAAAAAAAAWTTTTTTTGTATTTTATATATATAAAA 569

Qy 443 tcaaatTTTataataaaaaattgtgttaaataattacagttatatatttttttatctc 502  
||| ||| ||||| ||| : | ||| ||| ||||| |||  
pb 570 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaWMAATTTTCTTTCTTTCTTTCTTTCTTT 629

Qy 503 taattttatttgtcgccaatttttagttgatattttaacataaaaaaattgtacacat 562  
: ||||: || || ||||| || | || | : |||: || ||  
ph 620

Qy 563 ttacaagcccatatatacaataattatataaatattcattaaaaatatatttaaatatag 622  
||| ||||| ||| || ||||| : ||| || :|:| : |||||

Qy 623 gatataaatataactatattagaattattctactttaagataacatagggttaaatgtata 682  
| | |||: ||: | | || ||| : : | ||| : : | | :  
p1 755

Qy 683 attaataaggtagttattgttaagatgagtatatatgtcgtaaacataatcactaacc 742  
||||| ||||| :| :| :| ||: |||

Qy 743 atttttattaacttcttggtttgaagttccaaaaagaaatggaagggaatttgagag 802  
| | | : | : : | | : | : | | : | | | | | | |

Qy 803 taagttcatgtttatattacataatgaagttgatgtttcttcttttaaatattttta 862  
:||| | |||:| :| ::|| :| :| ||| | :| :|| | ||

Qy 863 tacaaaatatttaataaaataattaaggattgaatgaaaaatataatgaaagtcgtttt 922  
:: ||| | | | : | : | : | : | : | : | :

Qy 923 actaatagtcatt 937  
| | | : | | |

RESULT 7

LOCUS CNO503DQ 1101 bp DNA GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC  
BACR08109 of RPCI-98 library from Drosophila melanogaster (fru  
fly), genomic survey sequence.

|           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACCESSION | AL064580                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| VERSION   | AL064580.1 GI:4941932                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| KEYWORDS  | GSS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| SOURCE    | fruit fly.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| ORGANISM  | Drosophila melanogaster<br>Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;<br>Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;<br>Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| REFERENCE | 1 (bases 1 to 1101)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| AUTHORS   | Genoscope.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| TITLE     | Direct Submission                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| JOURNAL   | Submission (02-JUN-1999) Genoscope - Centre National de Sequencage :<br>BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr<br>- Web : www.genoscope.cns.fr)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| COMMENT   | Determination of this BAC-end sequence was carried out as part of a<br>collaboration with the Berkeley Drosophila Genome Project (BDGP).<br>The BDGP is constructing a physical map of the Drosophila<br>melanogaster genome using these BACs. For further information<br>please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> . The BDGP Drosophila<br>melanogaster BAC library was prepared by Kazutoyo Osoegawa and<br>Aaron Mammosser in Pieter de Jong's laboratory in the Department of<br>Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,<br>NY. The library is named RPCCI-98 and was constructed by partial<br>EcoRI digestion of Drosophila DNA provided by the BDGP from the<br>isogenic strain y2; cn bw sp, the same strain used for the BDGP's<br>P1 and EST libraries. A more detailed description of the library<br>and how to order individual BAC clones, the entire library, or<br>filters for hybridization from the BACPAC Resource Center can be<br>found at <a href="http://bacpac.med.buffalo.edu/drosophila/bac.htm">http://bacpac.med.buffalo.edu/drosophila/bac.htm</a> . |

```

FEATURES             Location/Qualifiers
  source              1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08109"
                     /note="end : TET3"

BASE COUNT           291 a      51 c      117 g      404 t      238 others
ORIGIN

```

Query Match 3.8%; Score 115.8; DB 122; Length 1101;  
Best Local Similarity 44.2%; Pred. No. 3.6e+08;  
Matches 253; Conservative 45; Mismatches 272; Indels 2; Gaps 1;

[illegible][illegible]

Qy 522 atttttagtgatattttaacataaaaaaaattgtacacatttacaagcccatatatacaaa 581  
||| : | : ||| : | : ||| : : | : | : ||| |||

Qy 582 taattatataaatatttcattaaaaaatatattttaaatataggatataaaatataactattt 641  
|| || || || || || || : : || || || : || || || || || || || || ||

Qy 642 tagaattattctactttaagataaacataggttaaagtataattaataagggttagtttat 701  
| : | : | | | | : : || : | | | | | | : : : | : |

Qy 702 tgtaagaatgagtatatatgtcgtaaacataatcactaaccatttttattaaactctctgg 761  
: | | | | | | : | | | | | | | | | |

Qy 762 ttttgaagtccaaaaagaaaatggaagggaaattgagagtaagttcatgtttatatta 821  
||| || | ||||| |||| | || | || | || | || |

Qy 822 tacataatgaagttgatgttttcttcttttaatatattttatatacaaaatatttaataaa 881

Qy 2346 aggtgataataatcttaatttgatgcagaggttaataatgggccgggtttgagctggactt 2405

```
117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result<br>No. | Query |       |        |     |          | Description         |
|---------------|-------|-------|--------|-----|----------|---------------------|
|               | Score | Match | Length | DB  | ID       |                     |
| c 1           | 133.6 | 4.4   | 1101   | 122 | CNS00EVL | AL069706 Drosophila |
| c 2           | 129.6 | 4.3   | 1101   | 122 | CNS0021J | AL061936 Drosophila |
| c 3           | 127.6 | 4.2   | 1101   | 122 | CNS00EVL | AL069706 Drosophila |
| c 4           | 120   | 3.9   | 804    | 120 | IL2681   | B12681 P27D1-Sp-1   |
| c 5           | 119.4 | 3.9   | 1101   | 122 | CNS0039G | AL063921 Drosophila |
| c 6           | 118.4 | 3.9   | 1101   | 122 | CNS003BD | AL064091 Drosophila |
| c 7           | 115.8 | 3.8   | 1101   | 122 | CNS003DQ | AL064580 Drosophila |
| c 8           | 115.2 | 3.8   | 1101   | 122 | CNS00B07 | AL069440 Drosophila |
| c 9           | 113.8 | 3.7   | 921    | 80  | AW727018 | AW727018 GA_Pa02    |
| c 10          | 109.6 | 3.6   | 1101   | 122 | CNS003BD | AL064091 Drosophila |
| c 11          | 108   | 3.5   | 1187   | 120 | IL1102   | B11102 P19C22-T7 J  |
| c 12          | 107.4 | 3.5   | 1101   | 122 | CNS00B07 | AL069440 Drosophila |
| c 13          | 107.2 | 3.5   | 890    | 93  | AQ026918 | AQ026918 CIT-HP-2   |
| c 14          | 105.6 | 3.5   | 1101   | 122 | CNS00B01 | AL057419 Drosophila |
| c 15          | 105   | 3.4   | 1101   | 122 | CNS00EPO | AL069493 Drosophila |
| c 16          | 104.8 | 3.4   | 1101   | 122 | CNS00L72 | AL078714 Drosophila |
| c 17          | 102.6 | 3.4   | 893    | 122 | CNS013KE | AL103436 Drosophila |
| c 18          | 102.6 | 3.4   | 1101   | 122 | CNS0039G | AL063921 Drosophila |
| c 19          | 101.4 | 3.3   | 1101   | 122 | CNS00KAE | AL077628 Drosophila |
| c 20          | 101   | 3.3   | 1187   | 120 | IL1102   | B11102 P19C22-T7 J  |
| c 21          | 100.8 | 3.3   | 1139   | 117 | AQ089753 | AQ089753 HS_1153 A  |
| c 22          | 100.6 | 3.3   | 1101   | 122 | CNS00FFG | AL071206 Drosophila |
| c 23          | 100.2 | 3.3   | 1225   | 123 | CNS0161D | AL106171 Drosophila |
| c 24          | 100   | 3.3   | 1101   | 122 | CNS0021J | AL061936 Drosophila |
| c 25          | 99.8  | 3.3   | 1201   | 123 | CNS0167M | AL106396 Drosophila |
| c 26          | 99.2  | 3.3   | 928    | 122 | CNS00DKY | AL071865 Drosophila |
| c 27          | 98.4  | 3.2   | 766    | 79  | AW683426 | AW683426 NF011G12L  |
| c 28          | 98    | 3.2   | 974    | 122 | CNS00ITT | AL075432 Drosophila |
| c 29          | 97.8  | 3.2   | 804    | 120 | IL2681   | B12681 P27D1-Sp-1   |
| c 30          | 97.4  | 3.2   | 836    | 122 | CNS01100 | AL095642 Drosophila |
| c 31          | 97.4  | 3.2   | 836    | 122 | CNS01100 | AL095642 Drosophila |
| c 32          | 97.4  | 3.2   | 935    | 120 | IL0881   | B10881 P24K6-Sp-1   |
| c 33          | 96.6  | 3.2   | 1101   | 122 | CNS00EQL | AL069526 Drosophila |
| c 34          | 95.6  | 3.1   | 990    | 122 | CNS00G01 | AL065624 Drosophila |
| c 35          | 95.6  | 3.1   | 1101   | 122 | CNS003BB | AL064089 Drosophila |
| c 36          | 95.2  | 3.1   | 963    | 122 | CNS00AAL | AL054918 Drosophila |
| c 37          | 95.2  | 3.1   | 1225   | 123 | CNS0161D | AL106171 Drosophila |
| c 38          | 95    | 3.1   | 893    | 122 | CNS013KE | AL103436 Drosophila |
| c 39          | 95    | 3.1   | 1101   | 122 | CNS003B4 | AL064082 Drosophila |
| c 40          | 94.8  | 3.1   | 1248   | 120 | IL1336   | B11336 P19M0-Sp6    |
| c 41          | 93.4  | 3.1   | 1006   | 122 | CNS00813 | AL051905 Drosophila |
| c 42          | 93.2  | 3.1   | 828    | 113 | AQ739398 | AQ739398 HS_5482 B  |
| c 43          | 93    | 3.1   | 966    | 122 | CNS0058C | AL061991 Drosophila |
| c 44          | 93    | 3.1   | 1101   | 123 | CNS017KE | AL108152 Drosophila |
| c 45          | 92.8  | 3.0   | 828    | 113 | AQ739398 | AQ739398 HS_5482 B  |

## ALIGNMENTS

```

RESULT 1
CNS00EVL/c
LOCUS      CNS00EVL      1101 bp      DNA      GSS      04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPC1-98 library from Drosophila melanogaster (fruit

```

fly), genomic survey sequence.

ACCESSION AL069706  
VERSION AL069706.1 GI:4949849  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Bukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

**AUTHORS** Genoscope

TITLE Direct Submission

**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/drosophila\\_melanogaster/BAClibrary.html](http://www.fruitfly.org/TheBDGP/drosophila_melanogaster/BAClibrary.html) was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

```
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1ib="RPCI-98"
/clone="BACR29B23"
/note="end : 77"
```

|            |       |      |      |       |            |
|------------|-------|------|------|-------|------------|
| BASE COUNT | 419 a | 91 c | 60 q | 299 t | 232 others |
|------------|-------|------|------|-------|------------|

ORIGIN

Query Match 4.4%; Score 133.6; DB 122; Length 1101;  
Best Local Similarity 41.5%; Pred. No. 9.5e-11;  
Matches 229; Conservative 100; Mismatches 219; Indels 4; Gaps 2;

Qy 2135 attttactaatattattataaaattgtagaatgattatntttcaataatttaacaa 2194  
|| :| | :| : ||| :| :| : | || | :| | : ||| :| | | ||

Db 1068 ATAANTAAATTWWTATATATWATWATATATTAWTTTWTAAAATATAWTTTAAATAA 1009

**Qy**    2195 caatattttaatatattattattattattttctcaatttttattaacaaaaacataaaatt 2254  
         |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 1008 ATATATWTAWTAWTATATATAWATAWAANTATAWAWTWWAATWAATWWATAWAWTTTAA 949

Qy 2255 ttgaca-aattaaataaatgaattaattttctcaatttttcgtgcaactattacaaaa 2313

Db 948 AANAATAWAWWWWTATATTTTTTTTTTTTTTWTWTATWTWTATATWTTWAAAAWA 889

Qy 2314 tccttcatagtcctaattttaatttgatgcagaggtgataataatcttaatttgatgcag 2373

Db 888 WWAANTWATATTTTTTTTATWATWATATTAWATATTTTTTTTWTATTTTWTATATWATWTA 829

Qy 2374 agqtaataatggccgggtttgaagctggacttaagcatgatattgacgtactttatattt 2433

Db 828 TWTWTWTTTWTATWTATWTTTATWTTWTTW--TATTWTATWTATATWTTWTTWTTT 772

Ov 2434 ttccaaattcaaccagctcaaatatgagctctaaaattttgtccaatttaatccaagcc 2493

Dh 771 A G A A T A T A T A T A T A T A T A T A T T W T W T W T T T A T T W T T T A A W T T A T A T A T A T T W 712

Ov 2404 gattttatcttcctccatatttttttatttttaaaaaatttatatgatttttttttaatt 2553

[illegible]

DD /11 TTTTWWTTTTTTTTTTTTTWRTWZZZATTTTTAAITAAAAAIIIAAAAIIIAAAI 032

6

-----

```

1  OPERATING SYSTEM:  PC-DOS/MS-DOS
2  SOFTWARE:  PatentIn Release #1.0, Version #1.25
3
4  CURRENT APPLICATION DATA:
5
6  APPLICATION NUMBER:  US/08/883,795A
7  FILING DATE:  27-JUN-1997
8  CLASSIFICATION:  435
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME:  Gravelle, Micheline
12 REGISTRATION NUMBER:  40,761
13 REFERENCE/DOCKET NUMBER:  7841-062
14
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE:  (416) 364-7311
17 TELEFAX:  (416) 361-1398
18
19 INFORMATION FOR SEQ ID NO: 36:
20
21 SEQUENCE CHARACTERISTICS:
22 LENGTH:  665 base pairs
23 TYPE:  nucleic acid
24 STRANDEDNESS:  single
25 TOPOLOGY:  linear
26
27 MOLECULE TYPE:  cDNA
28
29 ORIGINAL SOURCE:
30 ORGANISM:  Homo sapiens
31 IMMEDIATE SOURCE:
32 CLONE:  RH 32
33
34 US-08-883-795A-36

```

Query Match 2.5%; Score 75.4; DB 4; Length 665;  
Best Local Similarity 47.1%; Pred. No. 2.6e-05;  
Matches 296; Conservative 0; Mismatches 331; Indels 2; Gaps 2;

Yy 370 atcttcactcttgtaactcttgtaagtcaccaacttttaacaaagtctgttggtacata 429  
Db 648 ATTTTTTCCTCTCTACCTGGCAATATCAATCTCGAGATCAATTAACCTGCTCTTTTAA 589  
Yy 430 tatataatactctcaaaatttataataaaaaattgtgttaaatattacagtatatat 489  
Db 588 TTTTCTGCGCTCTTTTAACTATATTATAAAATATGAATATAAAATATGTAATATA- 530  
Yy 490 attttttatctctcaatttattttgtgcgcgaattttagtgtgattataacataaaa 549  
Db 529 ATACTTTAAATATAAAATATGTAATTATAAACTCTTAATATAAAATATGTAATTAA 470  
Yy 550 aaattgtacacattatacagccacatacaaaataattataaaattactataaaata 609  
Db 469 ATACTTTATAAAATATGTAATTATAAAATATGTAATTATAAACTTTTAAATATAAA 410  
Yy 610 tatttaaaataagtagataaaataacatttttagagattattcttacttaagtagaata 669  
Db 409 TGTAAATATAAACTTTTAAATATAAAATATGTAATTATAAACTTTTAAATATAAA 350  
Yy 670 ggttaagtgtataattacaagggttagttttatgttaagagtagtataatgtgttaac 729  
Db 349 TG-TAAATATAAACTTTTAAATATAAAATATGTAATTATAAACTTTTAAATATAAA 291  
Yy 730 ataatactacaccattttataactctgtgttttgtaagtcaccaagaagaagtgaag 789  
Db 290 ATTAAATATAAACTTTTAAATATAAAATATTTAAATATAAAATTTTAAATATAAA 231  
Yy 790 gaaattttgagagtaagttcagtttatattatacacaattgaagttggtgttctctct 849  
Db 230 ATTAAATATAAACTTTTAAATATAAAATATTTAAATATAAAATTTTAAATATAAA 171  
Yy 850 ttttaattttttatacaaaattttaataaaataattaaaggttgtagaataataata 909  
Db 170 ATTAAATATAAAATTTTAAATATAAAATATTTAAATATAAAATTTTAAATATAAA 111  
Yy 910 tgaaggtgtgtttactaatagctcattgtcattttgtgcactctactaaataatagata 969  
Db 110 ATTAAATATAAAATTTTAAATATAAAATATTTAAATATAAAATTTTAAATATAAA 51  
Yy 970 aattaaattggtgataagatgaagaaga 998

Db 50 TTTTAATTATAAAATATTTAATTATAAAA 22

RESULT 14

US -07-867-106-2/c  
Sequence 2, Application US/07867106  
Patent No. 5389526  
GENERAL INFORMATION:  
APPLICANT: Slade, Martin B  
APPLICANT: Chang, Andy C M  
APPLICANT: Williams, Keith L  
TITLE OF INVENTION: Improved Plasmid Vectors for  
TITLE OF INVENTION: Slime Moulds of the Genus Di  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: CS/07/867,106  
 FILING DATE: 19920625  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: AU PJ 7187  
 APPLICATION NUMBER: PCT/AU90/00530  
 FILING DATE: 02-NOV-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Peeney, Joanne Long  
 REGISTRATION NUMBER: 35,134  
 REFERENCE/DOCKET NUMBER: RICE-0002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-568-3100  
 TELEFAX: 215-568-3439  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5852 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ANTI-SENSE: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2378..5038  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2378..5038  
 CS-07-867-106-2

US-07-867-106-2

Query Match 2.4%; Score 74.4; DB 1; Length 5852;  
Best Local Similarity 48.3%; Pred. No. 4.9e-05;  
Matches 277; Conservative 0; Mismatches 286; Indels 11; Gaps 2;

|    |      |                                                            |      |
|----|------|------------------------------------------------------------|------|
| Qy | 406  | tacacgaattttgtgtgtgacatattatataatctttccaaattttatataaaattgt | 465  |
| Db | 5821 | TAATAAAGATCTTTTAAATTATATATACATTTTATAGTGTTATTAAATTATTTAT    | 5762 |
| Qy | 466  | gttttaaaattacacgattatattttttatctgaattttattgttcgcgaatttt    | 520  |
| Db | 5761 | TATGTGTGTTTGTGATTTTATATATAGCTATTTTGTGCTGTTTCTATATAT        | 5705 |
| Qy | 526  | tattgttgatttttccacgtataaaatgtgcacgtcttcacggcgcgatataatt    | 585  |
| Db | 5701 | CTATCTTTTATTTTAAATTTATATTTTAAATTTTAAATTTTAAATTTTAAATTTT    | 5642 |



QY 2183 ataatttatacaacattattataatatattattattatttccaaattttataaacc 2242

Db 1505 ATATTCGAATTATATTGTAAATTTTAAAAATATTAACTTTCTTATATGATCGATA 1446

QY 2243 aaaaactaaatttttgcacaaatkaaataagattattttcccaatttttcgtgcac 2302

Db 1445 ATCCATTAAGAAGGTATTACTATTTTGGAAATTTATCAAAAGTATATATTTTTTAC 1386

QY 2303 tattacaaaaactctcatagtcctcaacttaattttgatgcagaggcgataaatctct 2362

Db 1385 ATGAAAGATATATCTATTTTTTTAAATTATTAATTTTCTATAAATTTATATCTTT 1326

QY 2363 atttgatgcagaggtaataatggcggggtttgagctgacttaagcatgatattgaagt 2422

Db 1325 TATTGTATGATTCCTCATTTTAACTCACTACTATATAGAAAACATTATTAAGTT 1265

QY 2423 actttataatttttccaaattccaccaggctcgaaatagtgctaaatttgcacatt 2482

Db 1265 ACCAAATTATTTTATGATATATATAATTTTAAATTTACTAAATTTATTAATAAAT 1206

QY 2483 taatccagccomtttaagtgtgcataatttttttaattaaaaattatatcatc 2542

Db 1205 ATATGCTATTATATATCATGATCCACATTTAAAAATTTCTATGAAGTGGGTATATAT 1146

QY 2543 tttatttaatttaattattttataattttttattttatgtgaaatttttatatagtc 2602

Db 1145 ATTACT----TATATTGCTATTTTGGTATACGATATATCAAAATATGTTATTTTTA 1090

QY 2603 atcttaacattattgtaattgtttatattagagtagtatatatataattagtagtggtt 2662

Db 1089 ATTTTGTATAAAATTTAAATATAAATTTTAAATTTGAAATATAAACTTTTAAT 1030

QY 2663 atttttgtaataactaaaaatgggtctgtgggtgactgtgacatgaattgcctaa 2722

Db 1029 TTTCCT-GAAATATTATTAATAATTATATCATATATATATATGCAATTTCTCTAAA 971

QY 2723 actccaacttaattcatattttaaacaggctaatatttttttaacactgtttccaaatt 2782

Db 970 TTAACTAATTTTTTATAATATATATATATTAACACTATCTTCCTGTTATATATTTTT 911

QY 2783 ttctggcgtgaaattctctgcagctagatttaataaccacaggctcaatttgatgctca 2842

Db 910 TTTAATTATATATTTTAAATATTTTAAATTTATTAAGAAATTATACAGAAATTAGTAAATCCATT 851

QY 2843 atgaaatagaactcatattgagcttaattaaattccatt 2882

Db 850 TTGATATTTTTTTTTTTTTCATGATTAATTTTTTTTT 811

RESULT 11  
 US-08-544-332/s/c  
 : Sequence 8, Application US/08544332  
 : Patent No. 5935777  
 : GENERAL INFORMATION:  
 : APPLICANT: Moyer, Richard W.  
 : APPLICANT: Hall, Richard L.  
 : APPLICANT: Gruidl, Michael E.  
 : TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
 : NUMBER OF SEQUENCES: 77  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Gerard R. Bencon  
 : STREET: 2421 N.W. 41st Street, Suite A-1  
 : CITY: Gainesville  
 : STATE: FL  
 : COUNTRY: USA  
 : ZIP: 32606  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release 1.0.0, Version 1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/544,332

1 FILING DATE:  
 2 CLASSIFICATION: 435  
 3 PRIOR APPLICATION DATA:  
 4 APPLICATION NUMBER: US 07/991,867  
 5 FILING DATE: 07-DEC-1992  
 6 PRIOR APPLICATION DATA:  
 7 APPLICATION NUMBER: US 08/107,755  
 8 FILING DATE: 19-AGD-1993  
 9 PRIOR APPLICATION DATA:  
 10 APPLICATION NUMBER: WO 92/14818  
 11 FILING DATE: 12-FEB-1992  
 12 PRIOR APPLICATION DATA:  
 13 APPLICATION NUMBER: US 07/827,685  
 14 FILING DATE: 30-JAN-1992  
 15 PRIOR APPLICATION DATA:  
 16 APPLICATION NUMBER: US 07/657,584  
 17 FILING DATE: 19-FEB-1991  
 18 ATTORNEY/AGENT INFORMATION:  
 19 NAME: Bencen, Gerard H.  
 20 REGISTRATION NUMBER: 35,746  
 21 REFERENCE/DOCKET NUMBER: UFI14.C4  
 22 TELECOMMUNICATION INFORMATION:  
 23 TELEPHONE: 904-375-8100  
 24 TELEFAX: 904-372-5800  
 25 INFORMATION FOR SEQ ID NO: 8:  
 26 SEQUENCE CHARACTERISTICS:  
 27 LENGTH: 1511 base pairs  
 28 TYPE: nucleic acid  
 29 STRANDEDNESS: double  
 30 TOPOLOGY: unknown  
 31 MOLECULE TYPE: DNA (genomic)  
 32 ORIGINAL SOURCE:  
 33 ORGANISM: Amsacta moorei entomopoxivirus  
 34 FEATURE:  
 35 NAME/KEY: CDS  
 36 LOCATION: complement (18..218)  
 37 FEATURE:  
 38 NAME/KEY: CDS  
 39 LOCATION: complement (234..782)  
 40 FEATURE:  
 41 NAME/KEY: CDS  
 42 LOCATION: 852..1511  
 43

Query Match 2.5%; Score 76.4; DB 4; Length 1511;  
Best Local Similarity 46.3%; Pred. No. 1.9e-05;  
Matches 324: Conservative 0; Mismatches 371; Indels 5; Gaps 2;

Qy 2183 ataatatcaacacacattattatattatattatattattctccatttttatacaac 2242  
Db 1505 AATATCAATATATATTGGATAATTTTAAATATATATATTTCTTTATATGATGTA 1446  
Qy 2243 aasactaaaatttttgacaaattaaaaaaatgaattatttctcoatttttctgcaac 2302  
Db 1445 ATCAACATAAATGAGTTATATCTTTTGGGAATTTATCAAAATGATATATTTTTCATC 1386  
Qy 2303 tattacaaaaatccctcatgatctcaattcaatttttgagcagagggtataataactta 2362  
Db 1385 ATTGAAGATATATCTATTTTAAATTAATTAATTTTCTATAAATTAATATCTT 1326  
Qy 2363 atttgatgcagaggtaataatggccgggttgagctgactaatgatcatgatgaact 2422  
Db 1325 TATTGTATGATATGTCATATTTAACTCAATATCTAAATAGAGAAATATATTAAGT 1266  
Qy 2423 atttttgaatttcaaaataacacacagctgaaattactctgctcaaaattttgctaac 2482  
Db 1265 ACCAAATTATTTTACATATATTAATATTTTAAATAGCTAAATATTAATAAAT 1206  
Qy 2483 taatcagccaaacttttaagttgcacattatttttaatttaaaaaattatatctat 2542  
Db 1205 ATAGCATTTATATATCATGATGCACAAATTAATAATTTCTAGAGTGGGATATAAT 1146

Qy 2357 atcttaatttgatgcagaggttaataatggcgccgggttgagctggaactaagcatgat 2416  
| | | | |  
Db 1993 TTTTCTTTTCTTACTTTGAAAAAATAAAAAAAAAAAAAAAAAAACCCTCATTA 2052  
| | | | |  
Qy 2417 tgcagtaatttatattttccaaattccaccgcgcgaataatgatgttaaatattgt 2476  
| | | | |  
Db 2053 TAAATTAATTAATTAATTTGGTTTTTTTTGATTTTTTTTAAATAATTTAAAAATTATTC 2112  
| | | | |  
Qy 2477 ccaatttaaccagcccaatttaagtcgtccatatttttttaatttaaaaaattta 2536  
| | | | |  
Db 2113 TCTATCTANTTACTCTATTATAAATATGTGATATATATACAAATATTACAGTTT- 2171  
| | | | |  
Qy 2537 tatcatttttttttataatttttttttttttttttttttttttttttttttttttt 2596  
| | | | |  
Db 2172 TGCAATGACAAATTAAATATATTATTTTGGATTAATTTTTTTTTTTTTTTTTTTT 2231  
| | | | |  
Qy 2597 atagtcatttatacatttgtaagtgttttttttttttttttttttttttttttttt 2656  
| | | | |  
Db 2232 AAATTCCTTTTTTTTTTTTATTTTAAATTTTAAATTTTATTTTCCCACTTTCA 2291  
| | | | |  
Qy 2657 aggtttattttgttaataactaa 2682  
| | | | |  
Db 2292 TTTTATTATTATTATTATTGTAAA 2317  
| | | | |

## RESULT 8

US-08-451-405A-2/c

; Sequence 2, Application US/08451405A

; Patent No. 5736358

## ; GENERAL INFORMATION:

; APPLICANT: FASEL, NICOLAS JOSEPH

; APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE

; TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND

; TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THE WEBB LAW FIRM

; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE

; CITY: PITTSBURGH

; STATE: PENNSYLVANIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 15219-1818

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" FLOPPY DISK

; COMPUTER: Midwest Micro 486-50

; OPERATING SYSTEM: DOS

; SOFTWARE: WORDPERFECT 6.1

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/451,405A

; FILING DATE: 26-MAY-1995

; CLASSIFICATION: 435

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/965,273

; FILING DATE: 15-JAN-1993

; INFORMATION FOR SEQ ID NO: 2:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 731

; TYPE: NUCLEIC ACID

; STRANDEDNESS: SINGLE

; TOPOLOGY: UNKNOWN

US-08-451-405A-2

Query Match 2.6%; Score 78.4; DB 2; Length 731;  
Best Local Similarity 47.3%; Pred. No. 7.9e-06;  
Matches 299; Conservative 0; Mismatches 331; Indels 2; Gaps 2;

Qy 2149 attattttataaatttgtagatgtatttttttccaaatttaaacacaaatttttaatt 2208  
| | | | |  
Db 730 ATTTTCTTTAAATGTTTGTGTTAAATGTTTGTTTTTTTTTTAAAAAATAAATTT 671  
| | | | |  
Qy 2209 attattattattattttctcaattttttattaaacaaacacaaatttttgcaatttaa 2268  
| | | | |

Db 670 TGAATGATTA-AAGAAAAAATAAAAAATAATAGTAGAAAAAGGTATTTTTATT 612  
| | | | |  
Qy 2269 ataaatgaattattctcaattttttgtgcacattaccasaaattcttcagtctcta 2328  
| | | | |  
Db 611 AAAAAAGAAATATATATCTACTATTAGGAAAAATTTTTTTTCTAATGTATATATA 552  
| | | | |  
Qy 2329 atcttaatttgatgcagagtgataaattcttaatttttgatgcagaggttaaatggcc 2388  
| | | | |  
Db 551 AAATAAATAAATAAATATGTTATTTGTTGAATAAGGGGTGATGGTAAAAAATA 492  
| | | | |  
Qy 2389 ggggttgatgcaggttaagcatgatattgagctacttattttttccaaattccacc 2448  
| | | | |  
Db 491 AAAAAATAAAAAATAAAAAAATTTATTTAAAAAATAAATAAATAAAAAAATA 432  
| | | | |  
Qy 2449 agtcgaaatgatgctcaaaatttttcccaatttaaccagccatttaagtcgtc 2508  
| | | | |  
Db 431 AAATTTTAA-ACCTTTGAGATANTAGAGTGTGTTAAAGTGTGTTTGCACAAATAA 373  
| | | | |  
Qy 2509 catatttttttttaattaaaaaattatacatttttttaatttaatttttttt 2568  
| | | | |  
Db 372 GATATATAATATATTTTAAAAAATTTCTTAAACACTTTTTTTTTTGATTTTAA 313  
| | | | |  
Qy 2569 atattttttttttatgaaaatttttatatagtcatttaacattatgttaattttata 2628  
| | | | |  
Db 312 TTTTTTTTTTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTTGTTTCCA 253  
| | | | |  
Qy 2628 tttagtgatttatatatatttgatgatggtttattgtttaaaactaaatggg 2688  
| | | | |  
Db 252 ATCCAAAATATCTGAATTTTTTTTTTAGAATTTTCTTATCATATACCTGCACAACT 193  
| | | | |  
Qy 2689 tcttggtggtagatttgacatttaagtccaactcaacttaattcatttttaac 2748  
| | | | |  
Db 192 ATTTTGTAGTCACTATGTTTAAATAATTTTAAACAATAAAAACTCTTCAATAG 133  
| | | | |  
Qy 2749 aggccttaattttttattacacgtgttccaa 2780  
| | | | |  
Db 132 AAATATTATTCTAATTTTATTTTTTAAAA 101  
| | | | |

## RESULT 9

US-07-991-867B-8/c

; Sequence 8, Application US/07991867B

; Patent No. 5476781

## ; GENERAL INFORMATION:

; APPLICANT: Moyer, Richard W.

; APPLICANT: Hall, Richard L.

; APPLICANT: Gruidl, Michael E.

; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System

; NUMBER OF SEQUENCES: 66

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David R. Saliwanchik

; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/991,867B

; FILING DATE: 12-DEC-1992

; CLASSIFICATION: 435

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO 92/14618

; FILING DATE: 12-FEB-1992

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/827,685

; FILING DATE: 30-JAN-1992

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/657,584

REFERENCE/DOCKET NUMBER: UF114.C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: *Anaseta moorei* antemopovirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 852..1511  
US-08-544-352-8

Query Match 2.88; Score 84.2; DB 4; Length 1511;  
Best Local Similarity 43.64; Pred. No. 8.8e-07;  
Matches 573; Conservative 0; Mismatches 729; Indels 12; Gaps 4;

Qy 129 tctgttacaataataaatacatctgtagaataaattttattcaaatgaagctctaa 188  
Db 68 TTTTATTATATTTTGATATGTTTATTTAATCGTTTATGATATACAAATATTATTTA 127  
Qy 189 ccatctttaaattattgtatgttaattaaatgaagataaatacatctcttgagcatg 248  
Db 128 TCATTTTACCTATTTTTTTTCTATCTACTACGAAGATATCAGATTTCACCTTCAA 187  
Qy 249 tattttctatctaatgtttgtgctgttgatgattgtagtaagctgttttta 308  
Db 188 TATCAGATATAATATATCATATTTTGATTTATGATATAAATATATATATGATAT 247  
Qy 309 aatcacatatacatcttttgattgttatgataagtcacatacaagaatattgtgt 368  
Db 248 TATACATAATCTACACACAGCAATATAAATCTGTGCCACTTTTCAATTTTGTAT 307  
Qy 369 gatcttcaattttgaactttgataagtcacaaactttacaagaattgtgtgtacat 428  
Db 308 TTTTATTATGTTTATTAATCTGTAAGAGCACTTTTATACAAATATGATATAGCTT 367  
Qy 429 atatatatatatcttcaaattttataaataaattgttttaataattcaagttatat 488  
Db 368 GTAATTTTATTTTATTTTCTACTTATAGGATTAATTTTGATATAGATTT---AAATAT 424  
Qy 489 tattttttatctctaatttttattgtgtgcccattttttgttatatttaacataaaa 548  
Db 425 TCTGTGTAAGTACAAATTTATCCAGCACTAATCTTTTTTTTATTTATAGCACTT 484  
Qy 549 aaaattgtacacatttacaagcccatatacaaatatfatataaatattcattaaaaat 608  
Db 485 TATCAGAAATGTCTTAATCATTTTCTTCAAAAATGTACACTCATCTATGCCAATA 544  
Qy 609 atatttaaataggataaataataactattttagaattattctacttaagataaact 668  
Db 545 TATCATATTATCTACAGATATGTTTTCATTAATTAATTTTGTGTTTAAATGATAAAT 604  
Qy 669 aggttaattgataaataaagtttagttattgtaagatgagatatatgtctgtaaa 728  
Db 605 ATCTCTTATTAATATAATTCCTCATGATTTATTAATTTTATTAATAATCTATAT 664  
Qy 729 catatactactaacacatttttaataactctctgtgttttgagttccaaaagaataagaa 788  
Db 665 CTATATTATGAGTTAATATACACATTTTGTATAGATAAATAATATCTAATTTTTC 724

Qy 789 gggaaatttgagagtaagtctatgtttatattatatacaataagttgagttgtttc 847  
Db 725 GCATCAATCTGTGTTTGGCCAGAAACATAGACCAATATTATATCTATCGACATTT 784  
Qy 848 tttttaaattttttacaaatatttaataaataaataaagattgagtaaaaaaat 907  
Db 785 TTTTATTATTTTGATATATTTTTTCAAAAAAATATTATCAATCAAAAAAATATAAT 844  
Qy 908 aaatgaagctgttttactaatagtcattatgcattttgtgcactacttaataaaga 967  
Db 845 TATCAAAATGGATTACTAA-ATTCATGATATATTTAATAAATTTTAAATATTA- 901  
Qy 968 taatttaattgttggtacattagataaagaacaactagattttgcccacttctatgt 1027  
Db 902 ----TAAITAAAAAAATATATATAACAGAGATATCTTATTATATATATATTA 956  
Qy 1028 aaagctggtcgttacttaataaataaagttacattgacatgccacgtataactatctg 1087  
Db 957 AAAAAATAGTTAATTTAGAAGATTGCATATATATATATATATATATTTAAAT 1016  
Qy 1088 gtattctatcaactcagctaatttttaaagataaagattgtaatttttaataaga 1147  
Db 1017 ATATCTACGAAATATTAAAGTTTATATATTTCAATTAATATTTATTTAAAT 1076  
Qy 1148 aggtcaaatgtttttgtatcaacagctagggttaatttactattttcttaagaa 1207  
Db 1077 TTTTACGACAAATATAAATAATATACATATTATAGATATCTTTATACAAAATAGCAAT 1136  
Qy 1208 ataagtaataataaattgactttaaatacaaaactttctatgatacttttatcatatt 1267  
Db 1137 ATAAATATATATATATACCACTTCATAGAAATTTTAAATTTGGAATCATGATATA 1196  
Qy 1268 taactataatttaattgtgagagtaacaarataaaaaacatagaacaacccaagtt 1327  
Db 1197 AATGACTATATTTTATTAATTAATTTAGTAAATTAATAAATTAATATATCTAAAAAT 1256  
Qy 1328 agttatgtgtgactatatacagctaaataattgataaattttttctctgttacta 1387  
Db 1257 AAATTTGCTACTTTTAAATGTTTTCCTATATATAGTATAGTATTAATATGGAATCA 1316  
Qy 1388 attcctatcctgggtttttttttctagtagaacataaatttaaaaaataaata 1441  
Db 1317 ATACAAATAAAGATTATAAATTTATAGAAAAATTAATTAATTAATAAATAA 1370

## RESULT 6

US-08-883-795A-36

Sequence 36; Application US/0883795A

Patent No. 5985607

## GENERAL INFORMATION:

APPLICANT: Delcuve, Genevieve

APPLICANT: Awang, Gregor

TITLE OF INVENTION: Recombinant DNA Molecules and Expression

TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESS: BERESKIN &amp; PARR

STREET: 40 King Street West

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5H 3Y2

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/883,795A

FILING DATE: 27-JUN-1997

CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Gravelle, Michelle

Query Match 2.8%; Score 84.2; DB 1; Length 1511;  
Best Local Similarity 43.6%; Pred. No. 8.8e-07;  
Matches 573; Conservative 0; Mismatches 729; Indels 12; Gaps 4;

Y 129 ttgtgtacacataataaataacatcgatagaataaatttatctcaagtgaagcttaa 188  
Db 58 TTTTATTATTATTGTAATTTTGTATTAAATGCTATTGATATTACAAATTATTTA 127  
QY 189 ccaattctaatttttttagatgtatttaagaagaataaatcatctcttgagacgt 248  
Db 128 TCAATTACCTATTTTTTTTTTCTATTCTACTACAGAAATACGAGTTTGCCACTTCA 187  
QY 249 tattttcaacttaagtgtgtgtgctgtgtgtagtgatgtatgtagctgtcttta 308  
Db 188 TATCGAATAATAATTATCATATTTTTGCATTTATGAATAAAATATTAAATGATAT 247  
QY 309 aatccaatatcacacttttagtttggatgtatgtatagtcgaataaagatagtgt 367  
Db 248 TATACACATATCTACACAGGAACATATAATCTGTGTCACATTCTTCATTAATTGAT 307  
QY 369 gatcttcacattttgaacttggatagtcacaaacttttaacaaagtttgatgtacat 428  
Db 308 TTTTATTATGTTTTTAATGTGTAAAGAACACTTTTATACAAATTCACATATGCTT 367  
QY 429 atatatatatattctcaaaatttataataaaattgttttaaatatttcagttatat 488  
Db 368 GTATTTTTTTTATTTTCTCTCTTTAGGAATTATTTTGATATGAAATTTAAATAT 424  
QY 489 tattttttatctctaaattttatgttgcgaatttttagttgatttttaacataaa 548  
Db 425 TCTGTGTAAGTCACAAATTAAATCCGACCAACATACTTTTTTTTATTATAGCAATT 484  
QY 549 aaagtgtacacatttcacagccatacaataaattataaataattcattaaaaaat 608  
Db 485 TATCACAAATGTCTAAATCACTTTTCTCAAAAATTGACACTCATCTATGCCAATA 544  
QY 609 atatttaataaggataataaataaactattttagaattatctcatttaagaatacat 666  
Db 545 TATCATTAATTACTACGATATTGATTTCTAAATTAATTTTGTGTTTAATGTATAAT 604  
QY 669 aggttaaatgataaataaaggattgttttgtatagtaagatgatttatctgtgttaa 728  
Db 605 ATCTCTTATTATATATTTCGCGCATGATTTATATATTTTTTATTATAAATCTATAT 664  
QY 729 cataatcaatacaaccttttataaactcttggttttgaggtccaaaagaanaatgaa 788  
Db 665 CTATATTAGAGTTATATTACACATTTTGTATGATATAAATATATCTAATTTTTTC 724  
QY 789 gggaaatttggagtagtgatctgtattatataataatgaatgagttgtttc-ttc 847  
Db 725 GCATCATCTCTGTGTTTTGCCGAAACATAGGACCAATTATTAATCTATGACATTT 784  
QY 848 tttttatttttttatacaaaatttttaataaaataatgaagtgtgataaanaatat 907  
Db 785 TTTTATTATTATGATATATTTTTTCAAAAATAATTAATCAATGAAAAAATAAAT 844  
QY 908 aatgaagtgcttttataatgatcatgttcattgttgcgcatactaaatagaga 967  
Db 845 TATCAAAATGGATTACTAA-ATTCGTATATAATTTTAATAAATTTTAAATATTA-- 901  
QY 968 taaatttaattgttggtatagatacaagaacaaactagattgttccactctattgtt 1027  
Db 902 ----TAATTTTAAAAAAATAATAATACAGAGATATGTTATTATAATTATATATA 956  
QY 1028 aaagctgtctcgtttacataaaataggatcattgctatgcgcagctataactctgt 1087  
Db 957 AAAAAATTAGTAAATTAGAACAAATTCGATATATATAATTATGATATAAATTAATTAAT 1016  
QY 1088 gttattctatacaatcgcgaacttttataacagtgaagaatgatttttaaatagaa 1147  
Db 1017 AATATTCGAGAAATATAAAGGTTTATATATTTCAAAATTAATATATTATTAAT 1076

|    |      |                                                             |      |
|----|------|-------------------------------------------------------------|------|
| Qy | 1148 | agggtcaaatgtttatttgatcacaacagtaggattacttaacttattttccataagaa | 1207 |
| Db | 1077 | TTTATCAAAAATTTAAAAATATACACATTTAGATATCTTTATACAAAAATGCAAT     | 1136 |
| Qy | 1208 | ataagcaaatataatttgactcttaatacaaaaacttcctgatctttttatcatatt   | 1267 |
| Db | 1137 | ATAGATTAATATTATCTACACACTCTATAGAAATTTTAAATGGTGAACTGTAATATA   | 1196 |
| Qy | 1268 | tacttataatttatactgtgagatcaacaarttaaaaaacatagaaaacccaagaatt  | 1327 |
| Db | 1197 | AATGACATTAATTTTATTAATAATTAGTAAATTTAAAAAAATTAATAATATCTAAAAAT | 1256 |
| Qy | 1328 | agttatcgtgtgacatcacaacagctaaatttgaaataattttttttctgttaata    | 1387 |
| Db | 1257 | AAATTTGGTAATTATAATAGTTTTTCCATATTAGTATAGTGTAGCTAAATATGGAATCA | 1316 |
| Qy | 1388 | attccatcatcgggtttttttttttctagttaagccataattatacaaaataacaa    | 1441 |
| Db | 1317 | ATACAAATTAAGAGATTATAAAATATAGAGAAATTAATTAATTTAAAAAATTA       | 1370 |

RESULT 4

1 08-107-755a-8  
 2 Sequence 8, Application US/08107755a  
 3 Patent No. 5721352  
 4 GENERAL INFORMATION:  
 5 APPLICANT: Moyer, Richard W.  
 6 APPLICANT: Hall, Richard L.  
 7 APPLICANT: Gruidl, Michael E.  
 8 TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
 9 NUMBER OF SEQUENCES: 40  
 10 CORRESPONDENCE ADDRESS:  
 11 ADDRESSEE: David R. Saliwanchik  
 12 STREET: 2421 N.W. 41st Street, Suite A-1  
 13 CITY: Gainesville  
 14 STATE: Florida  
 15 COUNTRY: U.S.A.  
 16 ZIP: 32606  
 17 COMPUTER READABLE FORM:  
 18 MEDIUM TYPE: Floppy disk  
 19 COMPUTER: IBM PC compatible  
 20 OPERATING SYSTEM: PC-DOS/MS-DOS  
 21 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 22 CURRENT APPLICATION DATA:  
 23 APPLICATION NUMBER: US/08/107,755A  
 24 FILING DATE: 19-AUG-1993  
 25 CLASSIFICATION: 435  
 26 PRIOR APPLICATION DATA:  
 27 APPLICATION NUMBER: US 07/827,658  
 28 FILING DATE: 30-JAN-1992  
 29 PRIOR APPLICATION DATA:  
 30 APPLICATION NUMBER: US 07/657,584  
 31 FILING DATE: 19-FEB-1991  
 32 ATTORNEY/AGENT INFORMATION:  
 33 NAME: Saliwanchik, David R.  
 34 REGISTRATION NUMBER: 31,794  
 35 REFERENCE/DOCKET NUMBER: UF114.C2  
 36 TELECOMMUNICATION INFORMATION:  
 37 TELEPHONE: (904) 375-8100  
 38 TELEFAX: (904) 372-5800  
 39 INFORMATION FOR SEQ ID NO: 8:  
 40 SEQUENCE CHARACTERISTICS:  
 41 LENGTH: 1511 base pairs  
 42 TYPE: nucleic acid  
 43 STRANDEDNESS: double  
 44 TOPOLOGY: unknown  
 45 MOLECULE TYPE: DNA (genomic)  
 46 ORIGINAL SOURCE:  
 47 ORGANISM: Anaseta moorei entenopoxvirus  
 48 FEATURE:  
 49 NAME/KEY: CDS  
 50 LOCATION: complement (18..218)

[illegible]



Db 538 ATATTAAGCTATAGTCTTGGCTTGGCTATGTTCTAGAAAATTAATATATATAATATATA 479  
 Qy 1141 aa-----tagaagggtcgaattgttattgtctaacacgtaggattattacttat 1195  
 Db 478 TACTTTTAAATAGCTATAGTAAAAATATATATATCTAGTATATAAAGGCAAT 419  
 Qy 1196 ttctctaaagaataagtaaatataaattgcatcttaatacgaataattt 1246  
 Db 418 TTATTTAGAAGTTATGTAGATTAGTAAATATATATATAGATATACTT 368

RESULT 14  
 Q11710  
 ID Q11710 standard; DNA: 5852 BP.  
 AC Q11710;  
 DT 30-JUL-1991 (first entry)  
 DE Dictyostellum plasmid ddp2 containing Rep gene.  
 KW slime mould; replication; Rep gene; ss.  
 OS Dictyostellum discoideum.  
 FH Key Location/Qualifiers  
 FT cds 2378..5041  
 FT /\*tag= a  
 FT /product= involved in extrachromosomal replication  
 PN W09106644-A.  
 PD 15-MAY-1991.  
 PF 02-NOV-1990; A00530.  
 PR 02-NOV-1989; A0-007187.  
 PA (UTMA-) MACQUARIE UNIV.  
 PI Slade MB, Chang ACM, Williams KL;  
 DR WPI: 91-164194/72.  
 DR P-PSDB: R11988.  
 PT Polypeptide facilitating extra-chromosomal replication - of  
 PT recombinant plasmid in Dictyostellum species  
 PS Claim 15; Fig 1; 90pp; English.  
 CC The sequence of Ddp2 has been found to contain the putative open  
 CC reading frame indicated in the Features Table. The possible ORF is  
 CC flanked by regions with similarity to promoter and poly adenylation  
 CC signals of known Dictyostellum genes. The RNA and polypeptide  
 CC product of the Rep gene have not, however, been detected. It is  
 CC believed that the product is produced in low amounts to positively  
 CC regulate initiation of plasmid replication. The polypeptide may also  
 CC contain regions that act as negative regulators of plasmid copy  
 CC number. See also Q11711 and Q11712.  
 SQ Sequence 5852 BP; 2298 A; 651 C; 708 G; 2195 T;

Query Match 2.88; Score 84.4; DB 1; Length 5852;  
 Best Local Similarity 46.24; Pred. No. 0.0011;  
 Matches 317; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

Qy 1997 ttcttttaattcctttcttttacttactttataacgaattcttgatgaattgtccct 2056  
 Db 1635 TTTTCTTTGTCATGACACATCTTTTTTTTGTCTGACACATTTTTTAAAAA 1694  
 Qy 2057 acaaacatgtctatacaatgtttattataaattccattcttttactaagattt 2116  
 Db 1695 AAAAAAAGCTAAAACTATTGTCATGACATCAATTTTCTAGTTTTTTTATAGTGA 1754  
 Qy 2117 agtaactcgaactgctgatttttactaattattattataaattctagatgattt 2176  
 Db 1755 TATAAAAATTAATGCGCTATCGATATATACTTAATTTATAGATGAATATATTA 1814  
 Qy 2177 ttctctaaataatttaacaaataatttataattattattataaattcttacttttt 2236  
 Db 1815 TTTTAAATAAATCTACTTTTTTTTTTTCTCTTTTCTTTTAAATTTAAATTTT 1874  
 Qy 2237 taacaaacaaatcgaatttttgcacaaattaaataaagaattattcttcaattttt 2296  
 Db 1875 TTTTATTAGATCTCATAAATTAATAATCAATTTAAATTAAGATTTATTTAATATC 1934  
 Qy 2297 tgcactattcaaaactcctcttagtctcaactcttaattgtagcgaagggtataa 2356

Db 1935 CAAAACTATAAAAACTAATGTAGTTTAACCACTTTTCTATTCTTTTTTTTTT 1994  
 Qy 2357 atcttaattgttcgagaggttaataatggcgaggtttgagctgagcttaagcagat 2416  
 Db 1995 TTTTTTTTACTTTGAAAAA 2054  
 Qy 2417 tgcagctacttattatttttccaaattcaaccgcgagatagctcaaaatttgt 2476  
 Db 2055 ATATTAATCTACTGTGTTT---TTTGTGATTTTTTTAATAAATTAATAATTC 2111  
 Qy 2477 ccaatttaaccgaacccatttaagttgcacattatttttatttaaaatttta 2536  
 Db 2112 TCTATCTAATATATACCTATTATATAAATTAATGAATATATCAAAATATTATCAGT 2171  
 Qy 2537 tatcatttttttaataattattatttttatttttttttttttttttttttttt 2596  
 Db 2172 TGGCATGCAATTTTAAATATATATTTTTTGTATGATTTTTTTTTTTTTTTTTT 2231  
 Qy 2597 atagtcacttaacattatgttaattttattatagtagtatattattatttagtat 2656  
 Db 2232 AAATTCCTTTTTTTTTTTTTTTTATTTTAAATTTTAAATTTTATTTTCCCACTTCA 2291  
 Qy 2657 aggtttattttgttaataaactaaa 2682  
 Db 2292 TTTTATTTTATTTTATTTATTTGAAA 2317

RESULT 15  
 Q28302  
 ID Q28302 standard; DNA: 1511 BP.  
 AC Q28302;  
 DT 12-FEB-1993 (first entry)  
 DE AmEPV tk DNA.  
 KW Entomopoxvirus; thymidine kinase; non-essential; regulatory sequences;  
 KW vector; ss.  
 OS Annsacta moorei.  
 FH Key Location/Qualifiers  
 FT cds 852..1511  
 FT /\*tag= a  
 FT /label= ORF\_Q3  
 FT complement (234..782)  
 FT /\*tag= b  
 FT /label= ORF\_Q2  
 FT CDS complement (17..218)  
 FT /\*tag= c  
 FT /label= ORF\_Q1  
 FT promote 750..890  
 FT /\*tag= d  
 PN W09214818-A.  
 PD 03-SEP-1992.  
 PF 12-FEB-1992; U00855.  
 PR 19-FEB-1991; US-657584.  
 PR 30-JAN-1992; US-827685.  
 PA (UFL) UNIV FLORIDA.  
 PI Gruid ME, Hall RL, Moyer RW;  
 DR WPI: 92-316172/38.  
 DR P-PSDB: R29653-55.  
 PT New viral vectors and chimeric vaccines - comprise entomopoxvirus  
 PT expression system contg. spherocidin or thymidine kinase sequences  
 PS Disclosure; Fig 3; 110pp; English.  
 CC The sequence given is derived from the Entomopoxvirus, Annsacta moorei  
 CC (AmEPV) and contains the thymidine kinase (tk) DNA sequence. The open  
 CC reading frames indicated in the features table encode the tk protein  
 CC itself and also other structural or regulatory genes associated with  
 CC tk. The tk gene maps near the left end of the physical map of the  
 CC AmEPV genome. This gene is not highly related to any other  
 CC vertebrate poxvirus tk gene. Thymidine kinase is a non-essential  
 CC protein which makes it's gene desirable as a site for the insertion of  
 CC exogenous DNA.  
 SQ Sequence 1511 BP; 640 A; 128 C; 98 G; 645 T;

Query Match 2.88; Score 84.2; DB 1; Length 1511;

00 5805 ~~~~~ 5946



Qy 1237 caaaaactttcatgatacttttatcatattttacttataatttaataattgtgaqaataac 1296

Qy 783 atggaaggaatttgagaggaagttcattgtttatattatatacaatgaagttgatgttt 842  
Db 300 ATATCAAAATAAAGACCTTTTATATATATACGAAATATATTTCTTATTTTATGTTT 359  
Qy 843 tctctttttttatattttatatacaaaatttttaataaaat---aattagaattgaat 898  
Db 360 TCAAAATTTATAGACTTATATATTTATTTATGATACATTAACAAATAAAATATAT 419  
Qy 899 gaaaataataagaaggtgtttttactaatagctattgcatttggcgtactactta 958  
Db 420 GAGTATATATGTAATATATTTTATTTTACAGTTTATATGTTTATGACATATAT 479  
Qy 959 aataatagataaatttaattgtgtacatagatcaagaacaaactgaattgtccat 1018  
Db 480 GTGATAAATAAAATGATTAATATATATATATATATCTCTGTGAATTTATTAAT 539  
Qy 1019 tctattgttaaaagctggtcgtttacattaaataagctacgttgcacagat 1078  
Db 540 GGTATATATATATATATATATTTTATATATTTTGAATAAAATATTA--AAT 597  
Qy 1079 aactatctgttatttctatcaatccagctaattttaacagataaagtgaatttt 1138  
Db 598 AAAAATTTGTGTTGGTAAATCATATAAGTCAAGCTCATATTTATCTCATTA 657  
Qy 1139 taatagaagaaggtcaattgtttattgatctaacacagtaggatttaattactatttt 1198  
Db 658 AAATAGAAAGT--AAATATAATTTTACGACAGTACATATATATATGATATATTA 715  
Qy 1199 cctaagaagaatgaataataattttgaattcttaatacaaaactttctgatacttt 1258  
Db 716 AAAAAAATAAATAACACAT-ATATATATATATATATATATATGATATATAT 774  
Qy 1259 atcatattttactataaatttaattgtgagagctacaaraatttaaaacataga 1315  
Db 775 GTTTAAGTATGATGAATAAACAAGTTCATAGAGAAGAAATTAATAGGATAAA 831

## RESULT 6

T72882  
ID T72882 standard; cDNA; 19124 BP.  
AC T72882;  
DT 12-SEP-1997 (first entry)  
DE Plasmodium var-7 gene.  
KW DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;  
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;  
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;  
KW Plasmodium; ss.  
OS Plasmodium vivax.  
OS Plasmodium falciparum.  
FH Key Location/Qualifiers  
FT exon 7317..15139  
FT /\*tag= a  
FT /\*number= 1  
FT intron 15140..16205  
FT /\*tag= b  
FT /\*number= 1  
FT exon 16206..17552  
FT /\*tag= c  
FT /\*number= 2  
FT /\*note= "no stop codon given"  
PN W09640766-A2.  
PD 19-DEC-1996.  
PP 07-JUN-1996; W09508.  
PR 07-JUN-1995; US-487826.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;  
PI Wellens TE;  
DR WPI: 97-052231.05.  
DR p-PDSB: W22475.  
PT New malaria vaccines - contains cysteine-rich DBL family protein  
PT binding domains homologous domains of the Duffy and sialic acid  
PT binding proteins

PS Claim 4; Page 56-61; 96pp; English.

CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to  
CC the Duffy binding like (DBL) family of genes which have homology to the  
CC Duffy antigen binding protein (DABP) and sialic acid binding protein  
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var  
CC family of genes modulate cytoadherence and antigenic variation of  
CC Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding  
CC protein (DABP) are soluble proteins that appear in the culture  
CC supernatant after infected erythrocytes release merozoites. DABP and SAMP  
CC mediate the binding of merozoites and schizonts to the erythrocyte  
CC surface. These proteins are necessary for erythrocyte invasion by the  
CC parasite. This sequence can be used in the compositions of the invention.  
CC The compositions are for the treatment and prevention of malaria, and  
CC comprise either a nucleotide sequence or encoded polypeptide of the  
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of  
CC genes having homology with conserved regions of DABP and SAMP. The  
CC compositions are used for the treatment and prevention of malaria. They  
CC are also used in the preparation of vaccines for inducing a protective  
CC immune response in a mammal to Plasmodium merozoites (especially  
CC Plasmodium falciparum or Plasmodium vivax).  
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 3.28; Score 95.2; DB 1; Length 19124;  
Best Local Similarity 45.18; Pred. No. 3.2e-05;  
Matches 591; Conservative 0; Mismatches 709; Indels 10; Gaps 6;

Qy 138 aataataataacatcgtagaataaattttattcaaatgaagct---taaccctct 194  
Db 5418 AATACGTACATGTATTATAGAAATATAAGATTATATATATAGGATATAAATAT 5477  
Qy 195 ttaattttgttagatgaatttaattgaagaataacatattcttgacattgtatttt 254  
Db 5478 TTAATATATATTTTTTATGCAATTATGTTATATATATATATATATACGATTAGT 5537  
Qy 255 caatttaattttgtgctctgttggttagtgattgtatgacagctcttttaaatcag 314  
Db 5538 TTTTGAATAATTTTAATATCATATAATAATAAATAGTTAAATATAGTATTT 5597  
Qy 315 atatacatttttagttgtatgtagatagtcagcaatcaacgaatactgggtgatctt 374  
Db 5598 CATACAAATACTACTTATAAGTATATCAATATAATATATATATATATATTATG 5657  
Qy 375 caattttgaattgtatagtcacaaacttaacaaagttg---attgtgacatata 431  
Db 5658 TTTTGTATGGGTGTATATAAGGCTAAGTATATAGGGTGTTCATATATATTTATA 5717  
Qy 432 tatatatattcttcaattttataaaaaattgtgttaataatttcaagttattatt 491  
Db 5718 TGTGATAGATCATATATAGTAAAT-ATATTTATTGTATATAGTCTGTGTAAGATG 5776  
Qy 492 tttttttctctattttttttgtgccaatttttagttgatatttcaataaaaaa 551  
Db 5777 ATATGCATATACAGTTAAGGGTTATAGTTTTTTTTTTTTTTTGTACATATATAA 5836  
Qy 552 atgtcacattttacagccatatacaataaattataaattttcaaaaaatata 611  
Db 5837 AATGATATACATCATATGCAATATTACAGATATATTTGTATAAAT-ATATATATA 5895  
Qy 612 tttaataataggataaataaactttttagaattattctacttttaagaatacag 671  
Db 5896 TATATATATATAAGACATTAATAACTATCATAGGTAATAGTTTATATATATCAT 5955  
Qy 672 ttaattgtataaataaaggttagttttgttgaagaagttagtatgtgttaaacat 731  
Db 5956 TTTTATATATATTTTTTTTGTCTTACTCTCTGCTCTCTTTTGTCTATATATAT 6015  
Qy 732 aatcaactcaacttttttaactctgtgttttgaaggttcaaaaagaaggtgaag 791  
Db 6016 AACAAATATAAACAATATCAGTTATGGTAATAAATAAATTTATCTCATATATGCA 6075  
Qy 792 aaatttggagtagtctcattttatattatacaataaagaattgtgtgtttctcttt 851

SQ Sequence 1864 BP; 786 A; 210 C; 44 G; 732 T;

Query Match 3.68; Score 109.6; DB 1; Length 1864;  
 Best Local Similarity 44.74; Pred. No. 8.3e-07;  
 Matches 559; Conservative 0; Mismatches 674; Indels 18; Gaps 6;

Qy 4 gatgagaccacatttttaastagtaaacctaaccaatttttaataaagctgactct 63  
 Db 1608 GATTAATAAATAATATCTAGATAATTAATATTTAGAGAAATAGATATATTTCT 1549  
 Qy 64 atagacagagctttttatctctctctctttgtctctctctaggtctgacatgagaa 123  
 Db 1548 AATNTWTAGAGAGNTATCTATCTCTTAATAATCTTAATAATTTTATAGAGNT 1489  
 Qy 124 tttcttggtgtacataataataacacatgtagaataaattttattcaataagtg 183  
 Db 1488 TATATTNNNTAGAGATTTTTTATAGTTCTTAGTAATTTTATNTAGTTAGTAT 1429  
 Qy 184 cttacacacttttaatttttagatgtaatttaaatgaagaataacatctctctg 243  
 Db 1428 TATATTTATTAATATTAATAAAGTTTATGAAGAAGTTTAAAGTCAATATATTTAT 1369  
 Qy 244 acatgtattttctcttaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 303  
 Db 1368 ATTTTAAATTAATTAATTAATTTTATAGTATTAATTAATTAATTAATTAATTA 1309  
 Qy 304 tttta-aatcacatcatcattttgagttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 362  
 Db 1308 TACTATNTTAAATATTAGTTTATAGTAAATTTTATATATTTTAAAGTAAATTA 1249  
 Qy 363 tgggtgtatctctctctttgaactttgataagtcacccaacttttaacaaattgtgt 422  
 Db 1248 TGTGTANNNTAGTATATCTTATTTATAGTATTAATTAATTAATTAATTAATTA 1189  
 Qy 423 gtacatatatatatatctctcaacttttaataaaattgtgttaataaatttaacg 482  
 Db 1188 TAGCTAAGTGTAAATATTTAAATAAATTAATTAATTAATTAATTAATTAATTA 1129  
 Qy 483 ttatattattttt--tatctctaaattttttgtgtgtgtgtgtgtgtgtgtgtgt 540  
 Db 1128 ATTTTATTTATTAATAATCTAAATTAATTAATTAATTAATTAATTAATTAATTA 1069  
 Qy 541 acataaaaaattgtcacacattttacagccatatacaaatattataaattatcat 600  
 Db 1068 ATATAATAAATTTTTTAAAGTTATAGTNTAATTTTAAATTTTAAAGTTAAT 1010  
 Qy 601 taaaaatatatttaaatataggatataaataaactattttagaattattctacttaa 660  
 Db 1009 AATTATTAATTTTTTATTAATAATTTTATAATTTTATAGTAAAGTTGTTAA 950  
 Qy 661 gataacataggttaagtgataaataaaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 720  
 Db 945 TAATAAATAAATAATAGTACTTTATTA--TTATATATTAATAATATTTTATAG 892  
 Qy 721 gtctgaacataactcactaaccttttttaactctgtgtgtgtgtgtgtgtgtgtgt 780  
 Db 891 TAGTGTAGTTTATATAAAGAGCTTTAAATATTTCTTAATTAGATTAATATATA 832  
 Qy 781 aaatgagagagaaattgagagtagtctctctctctctctctctctctctctctct 840  
 Db 831 AATAGAGAGTATATAAATAATTAATTTAGTATTTAAATAGGATGATCTTNTATNT 772  
 Qy 841 tttctctcttttttaattttttacacaaatttttaataaataaattgaggttagtga 900  
 Db 771 ATAAAAATTAATAATTTTATATTAAGTACTAGTACTAGTATATNAGGCTAGAGAG 712  
 Qy 901 aaaaataaattgagagctgttttactactagtagtctgtgtgtgtgtgtgtgtgtgt 960  
 Db 711 TTTATATAATTAATAAATTTTATAATNTTAAAGAGATATATTTTTTATAGTAT 652  
 Qy 961 taatgataaatttaattgtgtacttagctcaagacaaactgattttgtcccaattc 1020

Db 651 ATAAGAGTTATTATTAAAGTAGCTATTTTCTTAG-----GATCTATTATATAN 599  
 Qy 1021 tattgttaaaagctgtctgtttacataaaataagctacagtgtctgtgtgtgtgtgt 1080  
 Db 598 NGTNTTATATATATAATAAATCTTAAGAGCTTAGTAAATTTAAATATATTATTA 539  
 Qy 1081 ctatctgtgtttctatcatcaacgtcaatttttaacagtagaattgagttgtgtgtgt 1140  
 Db 538 ATATTAAGGTATGTGNTGCTTGTGNTGTTTNTAGAAATATATATATATATATA 479  
 Qy 1141 aa-----tagaagggtcaattgtttgtatgtacacagctaggtgaatttacttat 1195  
 Db 478 TACTTTTTTATAGTATATAGTAAATTAATATATATATCTAGTATATAAGGNAAT 419  
 Qy 1196 tttcttaagaagaataaataataattggaacttatacaaaaacattt 1246  
 Db 418 TTTATCTAGAGGTATGTAGATNTAGTTTATATAATATTAGAATAGT 368

## RESULT 4

T72882/c

ID T72882 standard; cDNA; 19124 BP.

AC T72882;

DT 12-SEP-1997 (first entry)

DE Plasmodium var-7 gene.

KW DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;

KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;

KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;

KW Plasmodium; ss.

OS Plasmodium vivax.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT exon 7317..15139

FT /\*tag= a

FT /\*number= 1

FT intron 15140..16205

FT /\*tag= b

FT /\*number= 1

FT exon 16206..17552

FT /\*tag= c

FT /\*number= 2

FT /\*note= "no stop codon given"

PN W08640766-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; U09508.

PR 07-JUN-1995; US-487826.

PA (USSH ) US DEPT HEALTH &amp; HUMAN SERVICES.

PI Chitnis C, Millier LH, Peterson DS, Sim KL, Su X;

PI Wellens TE;

PW WPI; 97-052231/05.

DR P-PSDB; W22475.

PT New malaria vaccines - contains cysteine-rich DBL family protein

PT binding domains homologous domains of the Duffy and sialic acid

PT binding proteins

PS Claim 4; Page 56-61; 96pp; English.

CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to

CC the Duffy binding like (DBL) family of genes which have homology to the

CC Duffy antigen binding protein (DABP) and sialic acid binding protein

CC (SAMP) conserved regions (see T72889 and T72888 respectively). The var

CC family of genes modulate cytoadherence and antigenic variation of

CC Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding

CC protein (DABP) are soluble proteins that appear in the culture

CC supernatant after infected erythrocytes release merozoites. DABP and SAMP

CC mediate the binding of merozoites and schizonts to the erythrocyte

CC surface. These proteins are necessary for erythrocyte invasion by the

CC parasite. This sequence can be used in the compositions of the invention.

CC The compositions are for the treatment and prevention of malaria, and

CC comprise either a nucleotide sequence or encoded polypeptide of the

CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of

CC genes having homology with conserved regions of DABP and SAMP. The

CC compositions are used for the treatment and prevention of malaria. They

CC are also used in the preparation of vaccines for inducing a protective

CC immune response in a mammal to Plasmodium merozoites (especially

Db 181 AGCTTAAACCACTTTAATAATTTGAGATGAATTTAAATGAAGAATAATACATATCT 240  
Qy 241 tggacatgatatttcocttaattgttggtcctttgtagtaggttattgattacgat 300  
Db 241 TGGACATGATTTTCACTTAATGTTTGTGGCTTTGGTAGTGATTTGATGATGAT 300  
Qy 301 gtcttttaaacacacatacacttttgagttgttagatgataagtcgacataaagcaa 360  
Db 301 GTCTTTAAATACATATCACAATTTGAGTTTGTAGTAGATAGTGACATAAAGCAA 360  
Qy 361 tatgtgtgtgacttccattttgaaacttgataagtcacacaaactttgaaattgatt 420  
Db 361 TAGGTGTGACTTCACTTTTGAACCTTGATAAGTACCAAACTTTACAAAGTTGATT 420  
Qy 421 gtgtacatatatatattcttcaaatttataaasaatttgatttaaatattac 480  
Db 421 GTGTACATATATATATATCTCAAAATTTATAAATAAAATTTGTTTAAATAATTAC 480  
Qy 481 agttattattttttttatctctattttttgttgcacaaatttttagttgattttta 540  
Db 481 AGTTATTATTTTTTTTATCTTAATTTTATTGTGGCAAAATTTTAGTGATATTTA 540  
Qy 541 acataaasaaattgtacacatttacaagccatacaasataattataaattattcat 600  
Db 541 ACATAAASAAATTTGACACATTTACAAGCCATTAACAATAATTTATAAATTTGAT 600  
Qy 601 taasaaatattttaatatagataaataataacttttagaattattctacttaa 660  
Db 601 TAAAAAATATATTAAATATAGGATATAAATAAATCTTTTAAATATTCTCACTTAA 660  
Qy 661 gatacactagtttaagtataaataaagtttagttttgtttaagtagatattat 720  
Db 661 GATAACATAGTTTAAATGATATAAATAAGTTAGTTTGTAAAGTAGATATAT 720  
Qy 721 gtgttaaacataactaacacttttttaactcttggtttgaagttccaaaga 780  
Db 721 GTGTAAACATAACTAACATTTTATTACTCTGTGTTTGAAGTCCAAAGA 780  
Qy 781 aaatgaaaggaatttgagataagtcattttatattacataaagttgatt 840  
Db 781 AAATGAAAGGAATTTGAGATGAAGTTCATGTTTATATTATACATTAAGTGAATG 840  
Qy 841 ttctctctttttaattttttatacaaaattttaaataaataaaggtgaatga 900  
Db 841 TTCTCTCTTTTAAATATTTTATACAAATATTAAATAAATTAAGGATGAATGA 900  
Qy 901 aaaaataataaagtcgttttactaatagatattgttcattttgtgcataactaaa 960  
Db 901 AAAATATATAAGAGTGTCTTTACTTAAGTCAATTTGATGATTTGTGCGCATCTTAA 960  
Qy 961 taatagataaatttaattgtgtacttagatcaaaagcaaaactgaatttgcacatc 1020  
Db 961 TAATAGATAAATTAATGTGGTACATAGATCAAAAGCAAACTAGATTTTGTCCATTC 1020  
Qy 1021 tattgttaaaagctgttcgtttacattaaaatgaagtcattgttaactgcacataa 1080  
Db 1021 TATTGTTAAAGCTGTGCGTTTACATTAAGATAGGTACATGTTACAGCCACGTATA 1080  
Qy 1081 ctactgttattctatcaactacgctcaattttacacgtgaagtcatttactattttc 1140  
Db 1081 CTACTGTGTTTATCTCAATCAACGCTAATTTTAACTAGTAAGTAAGTAAATTTTA 1140  
Qy 1141 aatagaagagtcgaattgttattgtctcaacgctgaggtgatttactattttct 1200  
Db 1141 AATAGAAGGCTCAATGTTTATTTGATCTACACGTAGGATTAATTTACTATTTCCT 1200  
Qy 1201 taagaataaagtaaaataaatttgactcttaatacaaaactttctgactattttat 1260  
Db 1201 TAAGAATAAGTAAATATAATTTGATCTTATACAAAACCTTCTGATCTTTAT 1260  
Qy 1261 catatttacttaaaatttaattgtgaggtgaacaaatttaaaacatgaacaccc 1320

Db 1261 CATATTTTACTTATAATTTAATATTGTGAGCTACAAATTAJAAJACATGAACACC 1320  
Qy 1321 aaaaagttagttgtgtgtgactatatacacaggttaaaatttgaataattttttctc 1380  
Db 1321 AAAAGTTAGTATGGTGTGACTCATATACACAGTAAATTTGAATAAATTTTTCTTC 1380  
Qy 1381 gcttaataactctactgttttttttttttttctagtagaagcctaataatacaaatc 1440  
Db 1381 GTCAATAATTCATATGCGTTTTTTTTTCTAGTTAGCAATAATATCAAAATATTC 1440  
Qy 1441 atcataactctatcaactaccgcgctgctccctccctcaatacttaaacacataa 1500  
Db 1441 ATCAATAATCTATCAATACCCGCCCTGCTCCCTCCCTCAATATTAACCCACATA 1500  
Qy 1501 caccagcaccacacacacacttaataagccactatttttagcattctcttgcattaa 1560  
Db 1501 CACCAGCACCAACGCACTTAATAGCCACTTATTCAGCATGCTCTGCACTTAA 1560  
Qy 1561 gaaagtaagcttaactgcaattctcatatagcaggtctcaacagataaagttgtg 1620  
Db 1561 GAAAGTAAAGTACCTGCATCATTCATATGAGGCGTCAACAGTAAGTTGTGTG 1620  
Qy 1621 atgtgtgtgacacaggtttgaaacccgcccctcaactccctctttttctctctc 1680  
Db 1621 ATGGTTTGACACAGTTGTAAACCCGCGCTCCCACTCTCCCTTTCTTTCTCTCT 1680  
Qy 1681 coactcacaacctccaaattttctctatagttctatttaagttctttataacacag 1740  
Db 1681 CCATCCACACCCCTCAATTTCTCATATGTTCTATTATAGTCTTTTATAACACAG 1740  
Qy 1741 aatcagataagtcctcgaacacacacacacacgctctctgcagcaagctctgagtag 1800  
Db 1741 AATCAGATAGTCTCAGCAACACAAACACCATGCTCTGAGCAAGATCTGCACTAGT 1800  
Qy 1801 cagagctctgaatttgatcattattacagtcacaaacagatttaacaaagctgtgag 1860  
Db 1801 CAGAGCTCTGAATATGGATCATTTACAGTCAAAACAGTTTACAAAGCTCTCGAG 1860  
Qy 1861 ataaacactgaatctgctattgtttgtttttgtttacatattgttccagtgaaactatg 1920  
Db 1861 ATAAACACTGAATCTGCTATAGTTGTTTGTGTTTACATATGTTCCAGCTGAACTAG 1920  
Qy 1921 aagctctcttaagaacacacacacacacacacacacacacacacacacacacacac 1980  
Db 1921 AAGCATCTTAAGAAACCAACATCATATCAACCCATGATCAATGAATGATTTCA 1980  
Qy 1981 attttcgatgaatttcttttttcttttttttttttttttttttttttttttttttt 2040  
Db 1981 ATTTTCAGATTAAGTCTCTTTTAACTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2040  
Qy 2041 atgtgataatttccacacacacacacacacacacacacacacacacacacacacac 2100  
Db 2041 ATGTGATAGTCTCTTCAACACATGATCAATGTTTAAATTAATTTCTATCTCTCT 2100  
Qy 2101 attttactaagatttagtaacttcaactcgtgatttttactaattattattattataa 2160  
Db 2101 ATTTTACTAAGATTAGTAACCTCAACGCTGATTTTCTACTAATTATTATTATATA 2160  
Qy 2161 tttgtagatgatttttttctcaattttacacacacacacacacacacacacacacac 2220  
Db 2161 TTGTAGATGATATTTTCAATAATTTAACACAAATATTAAATATTATTATTATT 2220  
Qy 2221 attttcctatttttttaacacacacacacacacacacacacacacacacacacacac 2280  
Db 2221 ATTTCTCAATTTTATAACAAACATAAATTTTGAACAAATTAATTAATGATTA 2280  
Qy 2281 attttcctattttgtgacacttatacaaaactcctcctagctctcattattatttga 2340  
Db 2281 ATTTCTCAATTTTCTGCGCACTATACAAAACCTCTGATGCTCAATCTAATTTGA 2340  
Qy 2341 tgcagagtgataaattcttattttgtgagaggtataatggcggtttgtgctg 2400  
Db 2341 TGCAGAGTGATATATCTTATTGATGAGAGGTATATTGGCGGGTTGAGCTG 2400

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2206  
2207  
2208  
2209  
2210  
2211  
2212  
2213  
2214  
2215  
2216  
2217  
2218  
2219  
2220  
2221



Qy 411 aagttgtgtgtgtgcacatatatatatatctccaatttatataaaaattgtgttta 470  
Db 137610 AAGAGTGTATATATATATACCGACATATAATATATATACCAATATTTATTTGTTAT 137659

Qy 471 aaataattacagttatatattttttatctctcaatttttttgcgaatttttagt 530  
Db 137670 ATATATAAATCAAAATATTTTAAACATATATAAATAAATTTATATATTAATAT 137729

Qy 531 tgcatttttaacataaaaaaattgtcacacattacagcccatatacaaaatttat 590  
Db 137730 ATAAATAAATATTTATTTTAACTAATATATTTATATATATATTTAAATATAAAT 137789

Qy 591 aaatatctctaaaaaattatatataaatt--aggataaataaactatttaagat 647  
Db 137790 AATATTTTTTTTATTTATATATATACATATTAATTTATTAATAAATAAATTTTAAATATA 137849

AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Tamaki,T., Kurdi,O.B.,  
Conway,A.B. and Davis,R.W.  
TITLE Plasmodium falciparum 3D7 chromosome 12  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 130281)  
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.  
TITLE Direct Submission  
JOURNAL Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology  
Center, Stanford University, 855 California Avenue, Palo Alto, CA  
94304, USA  
COMMENT On Mar 15, 2000 this sequence version replaced gi:6652498.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 67262: contig of 67262 bp in length  
\* 67263 67462: gap of unknown length  
\* 67463 82485: contig of 15023 bp in length  
\* 82486 82685: gap of unknown length  
\* 82686 130281: contig of 47596 bp in length.  
FEATURES  
source Location/Qualifiers  
1..130281  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="12"  
/clone="3D7"  
BASE COUNT 52250 a 11780 c 11855 g 53996 t 400 others  
ORIGIN  
Query Match 5.5%; Score 168; DB 60; Length 130281;  
Best Local Similarity 44.2%; Pred. No. 3.8e-10;  
Matches 1146; Conservative 0; Mismatches 1416; Indels 29; Gaps 10;  
Qy 127 tcttggtgtacataataaataacatcgtagaataaatttattccaattgaagcttt 186  
Db 98689 TATTTCGTAAAAAATAATAAATAAGATATTAAATATTTATATTAACATCAAT 98748  
Qy 187 aacacatttaaatatttgtagatgaatttaaatgaagaataacatactatttggaca 246  
Db 98749 TAATATTTTATTTATATAAATATTATTATTAATAATATATAAATATTAATTATGGAA 98808  
Qy 247 tgtatttttcatttaattgttggtcttggtgatagggtattgatgtacagatgtctt 306  
Db 98809 TTTTAATTAATTAATTAATTAATTTATTTATTTATATATTTA-ATAAATAATATT 98867  
Qy 307 taatacocatcacacatttggattgtgatgatgaagtcagacataaagcaaatatgt 366  
Db 98868 TAAATTAATACACACAAAGATGCTTAAATTAATAATAAATATTTATTTATTTATAT 98927  
Qy 367 gtgatcttcacatttgaagcttgaatgcacccaacatttaacaagtttggatgtac 426  
Db 98928 GAATATAAATATTTAATTTAAAAATAAATAAATAAATAATATATTTATTTATAT 98987  
Qy 427 atatatatatatatctccaaatttataataaaattgtgtttaaataatttaccagtt 486  
Db 98988 A-ATAATTTAATTAATTAATTAATTTATTTAATTAATTAATTTAAAAATAAATAAAT 99046  
Qy 487 attattttttttatcttaattttatttgcgcgaatttttggatgatttttaacata 546  
Db 99047 AATATTATTATATATAATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 99106  
Qy 547 aaaaaaattgtacatttcaacgcccatacacaataattataaataattcttcaaaaa 606  
Db 99107 TAAATTAATTAATTAATTAATTAATTTATTTATTAATTAATTTAATTAATTAATTA 99166  
Qy 607 atatatatttaataaggatataataataacattttagaattattcttactttaagatac 666  
Db 99167 TATTATTATTA-ATTAAAAATAATAATAAATTAATTAATTAATTAATTAATTAATTA 99222

Qy 667 atagggttaagtataatttaaatagggtattgtttgttaagatgagtataatgtccta 726  
Db 99223 TAAAAATTAATTAATTAATTAATTAATTTATTAATTAATTTAATAAATTAATTA 99282  
Qy 727 aacataacatacaacatttttataactcttctgttttgaagtcacaaagaaagtg 786  
Db 99283 TAATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 99337  
Qy 787 aagggaatttgagagtaagttcagtttatattatacacaataagattgattttctt 846  
Db 99338 AATTATAAATAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 99397  
Qy 847 ctttttaattttttatacaaaattataaataaataaataagattgaatgaataa 906  
Db 99398 TTTCTTAATTATTATTATTAATTAATTAATTAATTAATTAATTTTATTAATAAATA 99457  
Qy 907 taatgaagattgtttactaatagtcattgtcattgttgcactactaaataatag 966  
Db 99458 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 99517  
Qy 967 ataatataatgttggtacattagatcaagaacaactgattgttgcctctctattgt 1026  
Db 99518 ATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 99577  
Qy 1027 taaagctggtcgtttactacataaataaggtacatgctacgcacgtataactatct 1086  
Db 99578 TATATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 99637  
Qy 1087 ggttattctcatcaatcagcgtatttttaacagtagaagatgaatttttaataaga 1146  
Db 99638 ATTAATATATAGTATAAATAATTTATGTATTATTAATTAATTAATTAATTAATTA 99697  
Qy 1147 aagggtcaattgttatttgatcacaacagtgaggatttaattacttattttctaaaga 1206  
Db 99698 TTTTATTATTATAAATAAATAATTTAATAAATAATTAATTTAATTAATTAATTAATTA 99757  
Qy 1207 aataagtaaaataaatttgaattcattacatacaaaactttcatgacttttcatcatt 1266  
Db 99758 AACATATAATTAATTAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTA 99817  
Qy 1267 ttaattataatttaatttggatgagtagtaacaarta-----aaaaacatgaaaccc 1320  
Db 99818 TCACATTTAATTAATAAATAAATAACCATTTATTAACTAATTAATTAATAAATA 99877  
Qy 1321 aaaaattgatttgggtgactcatatacacagattaaaattgaataaattttttcttc 1380  
Db 99878 AAAATAATTTAATGTGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 99937  
Qy 1381 gtcaatttccatcattggtttttttttttttagttagcacaattatcacaataatc 1440  
Db 99938 ATAAATTAATTAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTA 99997  
Qy 1441 atcataatccatcaatacccccgcctgcctccctccatcattcaaacccaactaa 1500  
Db 99998 TTTATTAATTAATTAATTAATTTATTAATTTAATTTGTTAATTAATTAATTAATTA 100057  
Qy 1501 caccagccacccacacacatttaagtcacacatttctgacgttgcctgacattaa 1560  
Db 100058 TAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 100117  
Qy 1561 gaaagtaagactaacctgcaatccatccatcgagggcccaacagataaagttggtg 1620  
Db 100118 ATATTATTAATTTAATTAATTTGTTTATTATTAATTAATTAATTAATTAATTAATTA 100175  
Qy 1621 atgggttgacacagattgttaaaacccggccctccatccctttttttctctctcc 1680  
Db 100176 TTTTATTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 100235  
Qy 1681 cccctccacccctcccaattttctctctgattgttctctctctctctctctctctcc 1740  
Db 100236 TAATTTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 100295



Qy 2054 cctacaacatgcoattacaagtgttaattataaattccattctcttatttactaagat 2113  
 Db 7745 ATTATATATTAATTTATATTAATTAATTTAAATTAATTTTAT-----1 7698

Qy 2114 attagtaactccaactgctgatttttactatttattattataaattgtagaagat 2173  
 Db 7697 ATTATTAATTAATGGTAAATTAATGAATATTATTATTAATTTATTGATTAAT 7638

Qy 2174 tatttttcoaatatttaacaacaattattattattattattatttttccatttt 2233  
 Db 7637 TATATATGATATATATATTAATTTATTTAAATTAATTAACATATATATTCTTTT 7578

Qy 2234 tatttaacaacaaactaaatttttgacaactaaataaactgaattattttccatttt 2293  
 Db 7577 TTATTAATTAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 7518

Qy 2294 togtgcaacttaacaaactctctcagctcattcttattttgtagcagagtgata 2353  
 Db 7517 ACAGTGTATATATATATATTAATTAATTTTATTTATTAATTAATTAATTAATTA 7458

Qy 2354 ataacttattttgtagcagagtgataaattggcgggtttgtagctggatttaagatga 2413  
 Db 7457 TATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7398

Qy 2414 tattgtagctattttatttttttccaaactaacaccgctgcaattatgcttcaaat 2473  
 Db 7397 TATATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7338

Qy 2474 tgtccaaatttaacaaagccatttttaagtgtgtctatttttttaatttttaaaat 2533  
 Db 7337 TGTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7278

Qy 2534 ttatatcatttttttttaatttaatttttttttttttttttttttttttttttttt 2590  
 Db 7277 GTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7218

Qy 2591 tttttatagctcttcaactatgttatttttttttttttttttttttttttttttttt 2650  
 Db 7217 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7158

Qy 2651 tagtataggtttttttttttaacttaaaactgggtttgtggctgagactggacc 2710  
 Db 7157 TTATATTAATTTTATTAATTAATTAATTAATTAATTTCTTTT-----ACTTCTAA 7105

Qy 2711 ttaactgtccaactcaacttaatttttttttttttttttttttttttttttttttt 2770  
 Db 7104 ATAAATATATGTAATTTCTTCAAGATATTATTTTCAATAATATATATTAATTA 7045

Qy 2771 ctgtttcaaaatttttcgggtgaa 2794  
 Db 7044 TAAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7021

RESULT 12  
 AC005504/c  
 LOCUS AC005504 104992 bp DNA HTG 01-APR-1999  
 DEFINITION Plasmodium falciparum chromosome 12, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 3 unordered pieces.  
 ACCESSION AC005504  
 VERSION AC005504.3 GI:4586584  
 KEYWORDS HTG: HTGS\_PRAEL.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 104992)  
 AUTHORS Hyman,R.W., Pung,E.L., Qin,F., Temaki,T., Kurd,I.O.B., Conway,A.B. and Davis,R.W.  
 TITLE Plasmodium falciparum 3D7 chromosome 12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 104992)  
 AUTHORS Hyman,R.W., Qin,F., Pung,E.L., Conway,A.B. and Davis,R.W.  
 TITLE Direct Submission

JOURNAL Submitted (21-APR-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 58642: contig of 58642 bp in length  
 \* 58643 58842: gap of unknown length  
 \* 58843 91011: contig of 32169 bp in length  
 \* 91012 91211: gap of unknown length  
 \* 91212 104992: contig of 13781 bp in length.

FEATURES  
 source Location/Qualifiers  
 1..104992  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5933"  
 /chromosome="12"

BASE COUNT 44286 a 9326 c 9564 g 4141 t 405 others

ORIGIN

Query Match 5.58; Score 168; DB 41; Length 104992;  
 Best Local Similarity 44.21; Pred. No. 4e-10;  
 Matches 1146; Conservative 0; Mismatches 1416; Indels 29; Gaps 10;

Qy 127 tctgtgttaccatataataacacatcagtagaataaatttttaccattgaagctct 186  
 Db 74983 TATTTCGTAAAAAAATATATAAAATGATTTAAATTTATTTATTAATTAATTAATTA 74924

Qy 187 aacactcttattattttgttagtgtaatttaaaagaagataacacattcttggaca 246  
 Db 74923 TAAATTTATTTATTTATTAATTTATTTATTTAAAAATATATAATTTATTTATTTGAAT 74864

Qy 247 tgtattttcatttattttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 306  
 Db 74863 TTTTAAATTAATTAATTAATTTGTTATTTAAATTAATTAATTAATTAATTAATTAATTA 74805

Qy 307 taactacacatcacattttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 366  
 Db 74804 TAAATTAATTAACACAAATGATCTTAATTAATTAATTAATTAATTTATTTATTAATTA 74745

Qy 367 gtgattctcattttttgaactttgataagcacaacttttaacaaagttgtgtgtgt 426  
 Db 74744 GAAATTAATTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 74685

Qy 427 atatatatatattcttcaattttttttttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 486  
 Db 74684 A-ATATATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 74626

Qy 487 atattttttttttcttcaattttttttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 546  
 Db 74625 AATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 74566

Qy 547 aaaaattgtcacacattttcaagccatatacaaatattataaattattcattaaaaa 606  
 Db 74565 TAAATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 74506

Qy 607 atatatataatttttttttttttttttttttttttttttttttttttttttttttt 666  
 Db 74505 TATTATTAAATTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 74450

Qy 667 ataggttaagtgtataatttaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 726  
 Db 74449 TTAATAATTAATTAATTTCAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTT 74390

Qy 727 aacataactcaactcaacttttttttttttttttttttttttttttttttttttttttt 786  
 Db 74389 TAAATTAATTAATTAATTTAAATTAATTAATTTAAATTAATTAATTTAAATTAATTAATTT 74335

Df 18676 ATATATATATAAATTTATTATATATATATATCTCATATAATTTAAATAAAAAATTTTA 18735

Qy 2577 tattttatgaaacttttatcatagcttcattcaacttagtgtaagtgtatatagaga 2636

Db 18736 AAATTTAAAAAGTGATGATATTAATTTAAAAAATTTTATATCTCATATTTATTATATTA 18795

Qy 2637 gtatttatatatattttagtataagggtttatttggtaaacctaasaaagggtcttgtgg 2696

Db 18796 ATTATTTTATATATAAATATATATATGAATTTAATTAATA--TATATATTATAAATTTA 18853

Qy 2697 gctagactgtgcaccttaaagctcgaacaactaatctatttttaaccagggtctaa 2756

Db 18854 TATATATTGAATATTTATATATATATATATATATATAGAAAATTTAAATTTTAAAT 18913

Qy 2757 tattttttatttaccactgttccaatttttcgggtgaaattcttcgatctagataasta 2816

Db 18914 AATTTTAATATAAATTTTTAAAAATTCCTCAAGGTATTTATTTTATAAAAAATTTTAT 18973

Qy 2817 acacccagggtctaatgttgctcacgaaatgaaattcatattgagcttaattaatat 2876

Db 18974 ATATAAAATCATTTTTTTTTAAAAATAACAAAAAATTTTTATAATAAATTTTATAA 19033

Qy 2877 tccattctctcttctgctga 2894

Db 19034 TGAATATATAATTTATTTA 19051

|            |                                                                                                                                                                                                                                                                                                                               |          |          |     |     |             |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------|-----|-----|-------------|
| AE001398/c | LOCUS                                                                                                                                                                                                                                                                                                                         | AE001398 | 14667 bp | DNA | INV | 06-NOV-1998 |
| DEFINITION | Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.                                                                                                                                                                                                                                                |          |          |     |     |             |
| ACCESSION  | AE001398 AE001362                                                                                                                                                                                                                                                                                                             |          |          |     |     |             |
| VERSION    | AE001398.1 GI:3845197                                                                                                                                                                                                                                                                                                         |          |          |     |     |             |
| KEYWORDS   |                                                                                                                                                                                                                                                                                                                               |          |          |     |     |             |
| SOURCE     | Malaria parasite P. falciparum.                                                                                                                                                                                                                                                                                               |          |          |     |     |             |
| ORGANISM   | Plasmodium falciparum                                                                                                                                                                                                                                                                                                         |          |          |     |     |             |
| REFERENCE  | Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium                                                                                                                                                                                                                                                                   |          |          |     |     |             |
| AUTHORS    | 1 (bases 1 to 14667)                                                                                                                                                                                                                                                                                                          |          |          |     |     |             |
|            | Gardner,M.J., Hethcote,H., Carucci,D.J., Cummings,L.M., Arayind,L., Koonin,E.V., Shaloom,S., Mason,T., Yu,K., Fujii,C., Pedersen,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Claytor,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L. |          |          |     |     |             |
| TITLE      | Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum                                                                                                                                                                                                                                                     |          |          |     |     |             |
| JOURNAL    | Science 282 (5391), 1126-1132 (1998)                                                                                                                                                                                                                                                                                          |          |          |     |     |             |
| MEDLINE    | E99021743                                                                                                                                                                                                                                                                                                                     |          |          |     |     |             |
| REMARK     | Erratum:[published erratum appears in Science 1998 Dec 4;282(5395):1827]]                                                                                                                                                                                                                                                     |          |          |     |     |             |
| REFERENCE  | 2 (bases 1 to 14667)                                                                                                                                                                                                                                                                                                          |          |          |     |     |             |
| AUTHORS    | Gardner,M.J.                                                                                                                                                                                                                                                                                                                  |          |          |     |     |             |
| TITLE      | Direct Submission                                                                                                                                                                                                                                                                                                             |          |          |     |     |             |
| JOURNAL    | Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA                                                                                                                                                                                                               |          |          |     |     |             |
| FEATURES   | Location/Qualifiers                                                                                                                                                                                                                                                                                                           |          |          |     |     |             |
| source     | 1..14667                                                                                                                                                                                                                                                                                                                      |          |          |     |     |             |
|            | /organism="Plasmodium falciparum"                                                                                                                                                                                                                                                                                             |          |          |     |     |             |
|            | /db_xref="taxon:5833"                                                                                                                                                                                                                                                                                                         |          |          |     |     |             |
|            | /chromosome="2"                                                                                                                                                                                                                                                                                                               |          |          |     |     |             |
| gene       | complement(1570..2424)                                                                                                                                                                                                                                                                                                        |          |          |     |     |             |
|            | /gene="PFB0490c"                                                                                                                                                                                                                                                                                                              |          |          |     |     |             |
| CDS        | complement(1570..2424)                                                                                                                                                                                                                                                                                                        |          |          |     |     |             |
|            | /gene="PFB0490c"                                                                                                                                                                                                                                                                                                              |          |          |     |     |             |
|            | /note="predicted by GlimmerM"                                                                                                                                                                                                                                                                                                 |          |          |     |     |             |
|            | /codon_start=1                                                                                                                                                                                                                                                                                                                |          |          |     |     |             |
|            | /product="hypothetical protein"                                                                                                                                                                                                                                                                                               |          |          |     |     |             |
|            | /protein_id="AACT1887.1"                                                                                                                                                                                                                                                                                                      |          |          |     |     |             |
|            | /db_xref="GI:3845198"                                                                                                                                                                                                                                                                                                         |          |          |     |     |             |
|            | /translation="MKEKNKIMDVLSCLPDDVVRKKSGKNSLKSSKTSYDYKK                                                                                                                                                                                                                                                                         |          |          |     |     |             |
|            | SSIFSKSKDHHKGGSGFRSGFINRKGSGFKPKYNNRLNNYNNYKGNFNFRNG                                                                                                                                                                                                                                                                          |          |          |     |     |             |
|            | DNFKTGKSGSGSGVSDPKNGKSGFKFARISINRNSKSSVYSTRGNSGNMGKSKNPAPSR                                                                                                                                                                                                                                                                   |          |          |     |     |             |
|            | TVYVTRVNTYKTSVDKPKNGKNNLISILKNNKFAITMKTKSGKPGVITIKSSVPPKRR"                                                                                                                                                                                                                                                                   |          |          |     |     |             |

REFERENCE 1 (bases 12511 to 12682)  
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Pauron,C.M. and  
Wolstenholme,D.R.  
TITLE Drosophila mitochondrial DNA: a novel gene order  
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)  
MEDLINE 83090428

REFERENCE 2 (bases 5269 to 5695)  
AUTHORS Clary,D.O., Wehleiter,J.A. and Wolstenholme,D.R.  
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5'  
flanking sequences and comparisons to mammalian mitochondrial tRNA  
genes  
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)  
MEDLINE 83220794

REFERENCE 3 (bases 404 to 5272)  
AUTHORS de Bruijn,M.H.  
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and  
genetic code  
JOURNAL Nature 304 (5923), 234-241 (1983)  
MEDLINE 83245048

REFERENCE 4 (bases 804 to 1778)  
AUTHORS Satta,Y., Ishida,H. and Chigusa,S.I.  
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in  
Drosophila melanogaster and its sibling species  
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)  
MEDLINE 88174373

REFERENCE 5 (bases 5268 to 13619)  
AUTHORS Garesse,R.  
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and  
evolutionary considerations  
JOURNAL Genetics 118 (4), 649-663 (1988)  
MEDLINE 88212147

REFERENCE 6 (bases 441 to 2967)  
AUTHORS Satta,Y. and Takahata,N.  
TITLE Evolution of Drosophila mitochondrial DNA and the history of the  
melanogaster subgroup  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)  
MEDLINE 91088557

REFERENCE 7 (bases 14215 to 14512)  
AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and  
Atkinson,P.W.  
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are  
modified arthropods  
JOURNAL Science 258 (5086), 1345-1348 (1992)  
MEDLINE 93088057

REFERENCE 8 (bases 14917 to 19517)  
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.  
TITLE Sequence, organization, and evolution of the A+T region of  
Drosophila melanogaster mitochondrial DNA  
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)  
MEDLINE 94285822

REFERENCE 9 (bases 1 to 408; 13319 to 19517)  
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
TITLE Drosophila melanogaster mitochondrial DNA: completion of the  
nucleotide sequence and evolutionary comparisons  
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)  
MEDLINE 96423163

REFERENCE 10 (bases 1 to 19517)  
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.  
TITLE Direct Submission  
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,  
Michigan State University, East Lansing, MI 48824-1319, USA

FEATURES  
source  
1. .19517  
/organism="Drosophila melanogaster"  
/organelle="mitochondrion"  
/db\_xref="taxon:7227"  
/note="derived from new and previously submitted  
sequences; sequence is a composite containing sequences  
obtained from different Drosophila melanogaster strains"  
1. .65  
/feature="tRNA-ND6"  
/product="tRNA-ile"  
/db\_xref="FlyBase:FBgn0013685"

tRNA  
1. .19517  
/gene="nt:ND6"  
/note="mitochondrial NADH-ubiquinone oxidoreductase chain  
6"  
/allele=""  
/db\_xref="FlyBase:FBgn0013685"  
complement(97..165)  
/product="tRNA-Gln"  
171..239  
/feature="tRNA-ND6"  
/product="tRNA-Phe"  
/db\_xref="FlyBase:FBgn0013685"  
240..1265  
/feature="tRNA-ND6"  
/codon\_start=1  
/db\_xref="FlyBase:FBgn0013685"  
/transl\_table=5  
/product="NADH dehydrogenase subunit 2"  
/protein\_id="AAC47811.1"  
/db\_xref="GI:1166532"  
/translation="MFNNSKSLPITIMIGTLITVTNSWLGANGLEINLSFIPL  
LSDNNLMSTASLAKTAVQLASTVLFPSSILLMLKXNNNNINSEFSMTINSALL  
LKSAGAPFPFPMNMGWETLWNAALMLTWQKATPLMLISLTKITLLISVILSVII  
GAIGGLNQTSLKLMFASFINELMGWLSLMSIESLWLLGFFPYSLSFLVTMFNFI  
KLHFLNQLFSWFWNSKILATFLFNLFLSGLGPPFGLPLFWLQQLTLCNFMFMT  
IMMSLTITLFFLRICYSAPMMYFENNMLKXNNNSINNTNMNITFISIFGLFII  
SLFTFMP"

CDS  
1264..1329  
/feature="tRNA-ND6"  
/product="tRNA-Trp"  
/db\_xref="FlyBase:FBgn0013685"  
complement(1322..1383)  
/product="tRNA-Cys"  
complement(1403..1468)  
/product="tRNA-Tyr"  
join(1470..1472,1474..3009)  
/codon\_start=1  
/exception="mechanism underlying reading frame shift after  
first codon uncertain"  
/transl\_table=5  
/product="cytochrome c oxidase subunit I"  
/protein\_id="AAC47812.2"  
/db\_xref="GI:7412849"  
/translation="MSRQLFSTNHRKDIGTLFIFGAWAGWGTSLILRAELHGPG  
ALIGDQIYNVIVTAFIMIFFWNPMIGFGNMVLPLMGAPNAPFRNNNSFW  
LLPPALLLLVSNVNGACGTGTVPPISAGIAHGASVOLAFSLHLGASSILGA  
VNFITTVINNRSTGISLDRLPLFVWSVITALLLLSLPVLGATINMLTDRLNLT  
FDPAGGSDPFLIQLHFWFHPVEVILLPGFMIHSIISQSGKXETFGSLGMIIYAN  
LAIGLGLFIWAAHMTVQMDVRYATFISATMIINAVPGIKFISLTLATGQLSYS  
PAILMALGFFFLFPGVGLGWSLVANSVDIILHDTITTVAFHFLVSLMGAVFALNAGF  
IEWPLFTGLTLNNKWLKSHFIIEIGVNLTFPFOHFLGAGHPRYSDFDATTWN  
IVSTIGSTISLGLIFFFPIINWLSVSRQVITPQIQLNNSIEWQVTPAERSTSELP  
LLTN"

tRNA  
3012..3077  
/feature="tRNA-ND6"  
/product="tRNA-Leu"  
/db\_xref="FlyBase:FBgn0013685"  
3083..3767  
/note="TAA stop codon is completed by the addition of 3' A  
residues to the mRNA"  
/codon\_start=1  
/transl\_except="(pos:3767,aa:TERM)  
/transl\_table=5  
/product="cytochrome c oxidase subunit II"  
/protein\_id="AAC47813.1"  
/db\_xref="GI:1166532"  
/translation="MSTWNLGLQDSAPLMELIIFDHALLILVIMTVLGVLMFM  
LFFNWNVRLGLQGLIEMWLLIPALITLALPSLWLLDEINFSVNTLSKIGH  
QWNVEYISDFNIEPDSIWTPIVELMOTGFLLDVQWNLVPMNSQILITLADVI  
HSWTVLGVKVDSTGPRNLQNTPIINRGLPTQCSICGASHSFMTIVISVPMY  
PIEWISSNNS"

CDS  
3768..3838  
/feature="tRNA-ND6"

repeat unit

|    |      |                                                                   |      |
|----|------|-------------------------------------------------------------------|------|
| Y  | 14   | aatttttaataatgaataaacctaaccaatttttataataaagaatgactccttagtaacaagag | 73   |
| Db | 1259 | ATAATATATATAAAATATTTTAAATATATAAAAATTTAAATATATTTTTATATAAA          | 1315 |
| Qy | 74   | cttttaatactctctctatttggctctctctcattgaggtggcaatcgagaattttctgtg     | 133  |
| Db | 1316 | MTCAATCATATATTTATATATATATATCATATATATTAATTTCAATTTCAATTAATATATA     | 1375 |
| Qy | 134  | tacacatataataatacatcgtgagaataaattttatccaattgaagctctaacacc         | 193  |
| Db | 1376 | GTATATAAAAAAATTATTTTTCATCACTAAATCGAATTAATTAATGTATATATAT           | 1435 |
| Qy | 194  | tttaattctgttagtgtaatttaastgaagaataaatcatctcttggcaatgtattt         | 253  |
| Db | 1436 | ATATATATATATATGTAATAAGAAATTAATTTATCCCTCATCAATTAATTTATATA          | 1495 |
| Qy | 254  | tcactctaaagt---tttgtctgggttgtaggtgattgatgtacagatgtctttaa          | 309  |
| Db | 1496 | TATTAATAACTTAAAAAAATATTTTTTTAAAAAAAATATTATTAAATTAAGTATA           | 1555 |
| Qy | 310  | atcacatactacattttgagtttgtatgatgaataagtcacataaangcaaatattggtg      | 369  |
| Db | 1556 | TAAACCATTTTTTAAATATAAATATTTTATAAATAAAATTTTAAATATTAATATAGA         | 1615 |
| Qy | 370  | atcttcactttgaaacttgtgaagtacgcaaaccaatttaacaagtttgattggtacata      | 429  |
| Db | 1616 | ATATTTTTATATATATAAAAAATAAAATATTTTTTAAAAAAATCAATTTATTTTATA         | 1675 |
| Qy | 430  | tatatatatctctcaaatttataataaaaatttgtttttaaataatttaccgttatatt       | 489  |
| Db | 1676 | TATATATATATATATAAATTTTAAATTTCAATTAATATATATAATATATAAATAAT          | 1735 |
| Qy | 490  | atttttttatctctaattttatttgtgcgaacatttttagttgatattttaacataaaa       | 549  |
| Db | 1736 | TTATTTTAATCACTAATCGAATAATTAATTTATATATATATATATATATAAAAAA           | 1795 |
| Qy | 550  | aaatgtgacacatttaacaagcgcatatacaaaattatataaatactcatttaaaaaata      | 609  |
| Db | 1796 | TGAAAATAAATTTATCCOCCATTCATAAATTTATGTATATATAAATCTAAAAATA           | 1855 |
| Qy | 610  | tatttaaatataggtataaataaactattttagatatactctactttaagaatacata        | 669  |
| Db | 1856 | TTTTTTTTTAAAAAAAAGATTTTAAATTTACTATAAAGTATTTTTATATATA              | 1915 |
| Qy | 670  | ggttaaatgataaatttaaatgagttgattttatgttaaaagagatataatctgtgtaac      | 729  |
| Db | 1916 | ATTATTTTATAAAATAAATATTTTAAATATTAATAAAAAATTTTAAATATAAATA           | 1975 |
| Qy | 730  | ataatcactaacccatttttataactctttgtggtttgaggttccaaaagaagtggaag       | 789  |
| Db | 1976 | ATTATAAGATTTTTTATAAAATCAATCATATATATATATATATATATATATATAT           | 2034 |
| Qy | 790  | gaaattttgagagtaagttctatgttatatatatacatgaattgaagttgattttctctt      | 849  |
| Db | 2035 | TTAATTTTCAATTAATTTATTAAGTATAAATAAAATTAATTTTATTTATCACTAAATCTG      | 2094 |
| Qy | 850  | tttaattttttatatacaaaattttaataaaaataaagaagttgagaaataataa           | 909  |
| Db | 2095 | AAATTAATTAATGTATATATATATATATATATATATATATGTAAGAAATGAATTAATTTAT     | 2154 |
| Qy | 910  | tgaagaatggttttaactatagtcataatgcaatttgcgcatacttaataataagata        | 969  |
| Db | 2155 | CCCCCATCTCAATAAATTTATTAATAATTAATAAACTTAAAAAAGATTTTTTTTTAAAAA      | 2214 |

```
misc_feature complement(39996. .40005)
      /gene="gap"
      /note="potential splice acceptor sequence"
misc_feature complement(40204. .40209)
      /gene="gap"
      /note="potential splice donor sequence, aag/gtaaca"
gene 45401. .50233
      /gene="MAL1P3.07"

very Match 5.8%; Score 175.8; DB 33; Length 67970;
1st Local Similarity 43.9%; Pred. No. 7e-11;
Matches 1220; Conservative 1; Mismatches 1539; Indels 22; Gaps 10;

139 atataataatacatcgtagaataaatttttcaaaattgaagcttaacaccttttaa 198
7824 ATAAATATATTCCTTCAATATATATATACAAATTTAAAAATATATATGAAGAAATAT 7883

139 tatttgtagatgtaatttaaatgaagaataacatattcttggacatgatttttcac 258
7684 TTCTTAAAAAACGATGATATATAATTAATTTAATATATATTAATCTTTATATATATA 7943

259 ttaattgttggtggttggtagaggtgattgtagcagatgctttttaaatocatat 318
7944 ATATATATATTTCCACATATTTTATATAGTTTATATATNA---ATTATATATNTCTTTAT 7999

319 caacttttggttgtagatgataagctgacataaangaaatttggtgtagcttccact 378
8000 TAATATATTAATATATAGTACACATTTATATTAATTAATATATTAAGAATTTATTTCT 8059

379 tttagcacttgtagaagtcaccaaaactttaacaaggtttgattgtagcatatataatat 438
8060 TTGTATATATATATATATATTAATATGGATAAATTAATAAATAATAAATAATAAATAA 8119

439 attcttcaaaattttataataaaatttggtttaaaataattcaagttatattatttttta 498
8120 TATATTAATAATATATAATTTACCAAAATATATATATATTAATTAATAAATAAATATAT 8179

499 tctctaattttttggctgcacaaattttttagtgatttttaacataaaaaaattgtcat 558
8180 TAAATTAATTAATTAATAAATAAATAATTAATAAATAATAAATTAATTAATTAATTAATA 8239

559 acatttaacaagcccatatcacaaatttatataaaattattctataaaaaaattttaaat 618
8240 TAAATATATTAATAATATTAATATTAATATATATTAATAAATAAATATATATTAATA 8299

619 ataggatataaaataaactattttagaagatttctactcttaagagaaacataggttaaatg 678
8300 AATTAAATTAATTAATTAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 8359

679 tataattataaagctgattttatgttaagaagtagtatatgtctgtgaaacataactact 738
8360 TATATTAATAAATAAATAAATAAATAATTAATAATTAATAAATTAATTAATTAATTAATTA 8419

739 aaocattttttataactcttgggttttgaaagtcocaaaagaaagaaaggggaatttg 798
8420 AAAATATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8477

799 agagtagtctgtgtttattattatcacataagaagttgatgtttttctcttttaaatatt 858
8478 ATATTAATAAATAAATAAATAATATATTAATAAATAATTAATTAATTAATAAATAAATAA 8537

859 tttatacaaaatttttaataaataaataaagagttgaatgaaaaataataagaagct 917
8538 ATTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8597

918 gttttactacagctcatcttgcatgttgctgcatctacttaataatagaataaataatt 977
8598 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8657

978 gtggctacttagatcaagaagcaactcagattgttgccactcttatttttaagaagctggt 1037
8658 TAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8717
```

Db 41588 AATTTTTTTTTTAAATATATATATATGTTAAATTTATCTAACCCATTAAATTT 41529  
Qy 182 gtcttaacacctcttaataatttgtagatgaatttaaatgaagataaatcatattctt 241  
Db 41528 TTTTAAAAATATATATATATATCTTTTAAATATCTGTAAATAAATGAATTTTTT 41469  
Qy 242 ggaacgtattttctcttaagcttttgcttggt---gataggtgtattgtagatgc 298  
Db 41468 TTTCAATTTAAATATTTATATATGTAATTTGTATATAAAGATATATATATAAATTAAT 41409  
Qy 299 atgtcttttaacacacacacacacacacacacacacacacacacacacacacacacac 358  
Db 41408 ATATATTTGTATATATATATATATATTAACCTTAATATAAACAATATACATATATAAA 41349  
Qy 359 aatatggtgtgatcttccttttgaacattgtagaagccacacacacacacacacacac 418  
Db 41348 TTAATTTAGCATATTTAATTTAAATTTATATTTAATTT-CAAAATATATATATATAATA 41290  
Qy 419 ttgtgacacataatataatcttccaaattttataataaaaaatttggtttaaataattt 478  
Db 41289 ATAAATCAGCATCAATA-AAATATATCTTTATTTAAATTTAAATTTAAATTTAAATTT 41231  
Qy 479 acagttatattattttttctctcaattttattgtgcgcacacacacacacacacacac 538  
Db 41230 TTA--TATATATATATTAATATATATAAAATCAATATAAAATTAATTTCAATAAAAATTT 41173  
Qy 539 taacataaaaaaattgtacacacacacacacacacacacacacacacacacacacacac 598  
Db 41172 AAATATACATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 41113  
Qy 599 attaaaaatatatttaaatataggatat--aaataaacatttttagaattattctactt 657  
Db 41112 TATATATATATATATATATATTAATATATTAATATATTAATTTAAATTTAAATTTAA 41053  
Qy 658 taagataacacacacacacacacacacacacacacacacacacacacacacacacacac 717  
Db 41052 ATTAATATACATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40993  
Qy 718 tatgtgcgaacacacacacacacacacacacacacacacacacacacacacacacacac 777  
Db 40992 ATATATATATATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40933  
Qy 778 agaaaaatggaaggaattttagagagtagtctgtttatataacataaagaattgga 837  
Db 40932 ATTAATATATATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40873  
Qy 838 tgtttctctcttttaattattttatacaaaaattttaaatataaataaataaagaattgga 897  
Db 40872 AATATATATATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40818  
Qy 898 tgaaaaataataagaagctgttttactaatagctatctgtcttctgtgcactcactt 957  
Db 40817 TTAATTTATATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40758  
Qy 958 aaataatagataaatttaattggtgacacttagctcaagaacacacacacacacacacac 1017  
Db 40757 TAATTTATATATATATATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40698  
Qy 1018 ttctattgttgaagaagctgtgcgttttacctataaataaagctacgttttcaatgcacgta 1077  
Db 40697 TTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40640  
Qy 1078 taactatctgtttattctatcaacacacacacacacacacacacacacacacacacacac 1137  
Db 40639 TATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40580  
Qy 1138 ttaaatagaagaagctgaattgttattgacacacacacacacacacacacacacacacacac 1197  
Db 40579 TTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40520  
Qy 1198 tctcaagaagaatagaataaataatttgatcttatacaaaaacacacacacacacacac 1257  
Db 40519 TAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40460

Qy 1258 tatctatttttacttaatttaattgttgagagtaacaaatttaaaaaacatagaac 1317  
Db 40459 TTAATTTAAATTTATATATATTTAAATTTATTTTAAATTTAAATTTAAATTTAAATTT 40400  
Qy 1318 accaaagaatttagtttggtgtgactatatacacagctaaatttgaataaatttttttc 1377  
Db 40399 AAGTTAAAAATATAAAATTAATTAATTAATTTAAATTTAAATTTAAATTTAAATTTATTA 40340  
Qy 1378 ttogtcttaattctcatcctgggtttttttttcttagttgaagcacaattatccaa-at 1436  
Db 40339 TTAATAATATATATATATATATATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAT 40280  
Qy 1437 aatcatcttaattctcatcacaaccccgccctgcctccctccctcaacttaaacocaa 1496  
Db 40279 TATAATATATATATATATATATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 40220  
Qy 1497 ctacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1556  
Db 40219 TTAATTAATCTGCTGTGTAATATATATATTTTAAATTTATTTAAATTTAAATTTAAATTT 40160  
Qy 1557 taagaagaagaagaacacacacacacacacacacacacacacacacacacacacacac 1615  
Db 40159 AATTAAAAATATAAAATGTTTCATAAAATAAATTTAAATTTAAATTTAAATTTAAATTTAT 40100  
Qy 1616 ggtgtgaggtttgacacacacacacacacacacacacacacacacacacacacacac 1675  
Db 40099 ATAAATTAATTTATTTATATATATATTTAAATTTAAATTTAAATTTAAATTTAAATTT 40040  
Qy 1676 cctcccccactccacacacacacacacacacacacacacacacacacacacacacacac 1735  
Db 40039 TATTATATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 39983  
Qy 1736 cacagaatacagaagaagctcgcacacacacacacacacacacacacacacacacacacac 1795  
Db 39982 CATTTAACTTTTATTTAAATTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTT 39923  
Qy 1796 ctatgcagagctcgaatttttgatcattatcagctcaaaacacacacacacacacacac 1855  
Db 39922 TGAATTTAA--ATGTTATATATATATTTAAATTTAAATTTAAATTTAAATTTAAATTTATTT 39864  
Qy 1856 tgacagataaacacacacacacacacacacacacacacacacacacacacacacacacac 1915  
Db 39863 TTAATATATTTATTTAAATTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 39804  
Qy 1916 ctatgaagctctctcaagaacacacacacacacacacacacacacacacacacacacacac 1975  
Db 39803 AAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 39744  
Qy 1976 ttcaatttttcgagtagaagcttcttttaattctctcttttttacttctttataacga 2035  
Db 39743 ATTAATATATATATATATTTAAATTTAAATTTAAATTTATTTATTTATTTATTTATTTATTT 39684  
Qy 2036 attctatggaataagcttctcacaacacacacacacacacacacacacacacacacacacac 2095  
Db 39683 AAAAAACATTTATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTATTTATTT 39624  
Qy 2096 ctctctattttcagaatattgtaactcacaacacacacacacacacacacacacacacacac 2155  
Db 39623 TTGAATTTAAATTTATTTATTTATTTATTTATTTAAATTTAAATTTAAATTTAAATTTATTT 39564  
Qy 2156 ataaattttctgaatttttttctcaacacacacacacacacacacacacacacacacacacac 2215  
Db 39563 TTAATTTATTTA---TTTATTTATTTATTTAAATTTAAATTTAAATTTATTTATTTATTTATTT 39508  
Qy 2216 ttattttttctcaatttttttcaacacacacacacacacacacacacacacacacacacacac 2275  
Db 39507 TAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 39448  
Qy 2276 ataaattttctcaatttttttgcacacacacacacacacacacacacacacacacacacacac 2335  
Db 39447 AAATATATATATTTATTTATTTATTTATTTAAATTTAAATTTATTTATTTATTTATTTATTT 39388

Db 9105 TCGTTTATTATTAATATTTATATATATATTTTATTTATTTAAATATATATTA 9046

Qy 1712 gtctattataagttcttattataacacagaatcagatagctctcgaacacacacac 1771

Db 9045 ATATAAGACATATATTAACATATAAATATATATATATTTATTTATTTATAT 8986

Qy 1772 catggtctcagacagatctgactagctagctcgaatcttgatctattacag 1831

Db 8985 TTGTTAATATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTA 8926

Qy 1832 tccaaacacgtaacaaagctgctgcagataaacactgaatctgctatagttgtttt 1891

Db 8925 TTATTTAATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8866

Qy 1892 ggtttt-----acatatgttccacgtaaacactgaagctctctcagaacacacac 1945

Db 8865 TATTTAATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8806

Qy 1946 atcatcatcacccctcgatcgaatcgaatcttgatcttgatcttgatcttgatct 2000

Db 8805 TTCTATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8746

Qy 2001 ttttaactcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 2060

Db 8745 TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8686

Qy 2061 acatgtctacacatgttttaattataaactcattctcttcttcttcttcttcttct 2120

Db 8685 TATTTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8626

Qy 2121 acttcaacactgctgatttttactatatttatttataaagtcttagaagatattttt 2180

Db 8625 ATTATTTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8566

Qy 2181 caataatttaacacacatatttaattattattattattattctcaatttttataaa 2240

Db 8565 ATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8506

Qy 2241 caaaacacataattttgcacaaataaataaagtgaattttctcattttcttctgca 2300

Db 8505 TATATTAATAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8446

Qy 2301 actattacacaaactctctatgctcctaacttcttcttctgacagaggtataaact 2360

Db 8445 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8386

Qy 2361 taatttgatcgagaggaataaagtggcggttgatgctgactta-----agcatgat 2416

Db 8385 ATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8326

Qy 2417 tgacgtactcttattttttcacaacacacgctcgaataactgactcaaaattttgt 2476

Db 8325 TATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8266

Qy 2477 ccaatttaacacagccatttaagcttgcataatttttatttttaaaanaattta 2536

Db 8265 ATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8206

Qy 2537 tatcattttatttttaattattattttattttattttattttatttttttttttt 2596

Db 8205 ATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8146

Qy 2597 atagctctcttaacatctgcttaagtcttattatagagtagatctatatatttagat 2656

Db 8145 TTGCTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8086

Qy 2657 aggtttattttgttaataaactcaaaatgggtcttggctgagactggacattaa 2716

Db 8085 A-TTTAATATATATATATATATATCAAGAGTAATAATATCTTAAATATTTATTTA 8026

Qy 2717 gctcaactcacaacttaactctatatttaaacagagcttaattttttattacactgtt 2776

Db 8027 TATAATGCTGACTATATTTATTTATTTATTTAAGAAATATATATTTATATTAACAT 7968

Qy 2777 caaatttttgggtgaattctctcagctctgattataataacacacacaggtctatgta 2836

Db 7967 TAAATTAATGCTGAATATATTTATTTATTTATTTAATTAATTAATATATTTATTT 7908

Qy 2837 tgctcaatga 2846

Db 7907 TATTTATATCA 7898

RESULT 7  
PFMAL3P5/c  
LOCUS PFMAL3P5 86829 bp DNA INV 11-FEB-2000  
DEFINITION Plasmodium falciparum MAL3P5, complete sequence.  
ACCESSION AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162  
AL010206 AL010210 AL139179  
AL034556.2 GI:4493931  
VERSION  
KEYWORDS HTG.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 86829)  
AUTHORS Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, R., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Ryes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Wewble, D.C. and Barrell, B.G.  
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum  
JOURNAL Nature 400 (6744), 532-538 (1999)  
MEDLINE 99376085  
REFERENCE 2 (bases 1 to 86829)  
AUTHORS Bowman, S., Skelton, J., Churcher, C., Lawson, D., Quail, M. and Barrell, B.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 86829)  
AUTHORS Lawson, D., Bowman, S. and Barrell, B.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK  
COMMENT On Mar 24, 1999 this sequence version replaced gi:4034877.  
For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).  
FEATURES  
source Location/Qualifiers  
1..86829  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="3"  
/clone="MAL3P5"  
gene 324..2944  
/gene="MAL3P5.1"  
join(324..668,1199..1303,1460..2944)  
/gene="MAL3P5.1"  
/note="predicted using hexExon; MAL3P5.1 (PF00575w), Hypothetical protein, len: 645 aa"  
/codon\_start=1  
/protein\_id="CAB38969.1"  
/db\_xref="GI:4493933"  
/db\_xref="SPTRMBL:097258"  
/translation="MKLVNVIYISSGFLPDLFCFSFHLIMKYKNNMKNVSVFEL  
RSQVIAKKRKRBRKINVSFKKKQKPLFLPNLAKGKSPGLGWRNQDVNKKKK  
KKKKKKKKKKPKVSLINQISSEKVKKKDAENYALHLPLADENTLFSMGMIDPF  
FKSQKVFICTRDIKSKRKKKNLSYINFLPCTIYPTTCWMLLKIVLSHLCIFFF  
PFPCPLCNILIRIYFSCVDLIRKKIRBYNLYCEKKRIKFKMDAKIKMKIDKDD  
LIFWYHDELCFCFMYKINIRNKKNIIRSNVDNNIDIDDMWNNPINDVNIKI  
SLDEIKIRQFENPDENLKSIDYFQFOLPNDNIYIREDQPLINDNNSNDIN  
NINWIKNNHKIKIDDDVDYKREDIYQKNDIDYTKITQNSLKEPKNQ  
IDQADIEQWFLSNWLDGRVRSNDEYKSETHIKNNKTKNGCTELIQNMNN





```

NIDYENNIYSIKSKFATPPITIKVILLOFTLLSVLHMKNWYKLLIDPDKIKFK
GETWLPPIPIIIGLITQILMFWNLIMSSVESHKPEFLIPIFETISL
PTIWHMNIENKIMVNLINENKFFIDCVSKTSRSTNNNNNNSSSSSSSSSSSS
HNNKSNIDIKEDQVHGLTSTVSTIITWRSRINBGLINCELPFAITVPTIPI
QNTLLNEFSLTIVASITFSTIITFKTFKFNALFKTILAVIITPLDAPQNM
KILLISILKKNGLPPLFQNCYIHNVPINMKNKNSOLKKNITPCNMWKNVNY
WSNPLIITILKPKNEIKMVPMLIKALLSKYIKTEKVLITCQIDREHITFRTR
DLTLLILILAKKNFTNTYIITLISKSILMKNKNMSTADALILITLSKNITLTDQI
YKNEIFWISLAFENHNLNMLNLSLSTFSTYSTYINNCFTITTTTFSFPIITK
PIIILKNLALNSTDLMPYIGSLIHLNMTIPIHVDQKPEKPNKKNENYNDHSH
ITLKNINNNCTDHRLLSDNFTIMTNDQSTYKIKQNNKIKQNNKIKQNNKIKQ
IQLIGTCTSSNNQQLIQIDQNEKNLLITDNRKNLKDPCQNVKDKGKKQ
NLIKNLIINIKIIEBKLSSTKQIENVLFSVLNNIINIKIIFPLNQKINIRNY
INITYWIKIYLNDEBENTSHCILKKNKKDILYHDMKFLNLMNEIYRDLN
MKQIILLGLGFLNNFQPEKILKRFILPKKEIKQIKGEMFLYQFPFVRLC
FNLKQKQMLFISQDEYENYIISDKYESAKLNNSTNMPSSLAKEKNTREHGDKWTL
DLTIRDDFLYMKNRKORYKIYLDNFNIPIATVVBQKDHIDYENNNFDMKS
FIQLKKNKINNNNNNNNNNNNIYIDITNQVKNKSTHNVIKNETNDNTPNS
TIRNQHPQDVLNVPFNNKLNVDSD:DFEYETCTNLNLDITIKVILKALLK
NKLKSLQSLIDLCIYEKLNIDRTIKYILNKLKELLIDNTYLLKVINIYKFLY
NMISTYQIKNLILPYNININISQILKILGMLSYHKNKNSPFBKNNVQNAARY
LNLNLIQNLIDQLKIEEMVNDLITFPTKLPKLSILNIVKTVQNCSTNKTZNM
THTNMLKNKQKQIIEHNDHGRHNNNNHNDVSSSSSSSTYLLKSGKGLG
NINVLNDIDINKISISYIKIQQIDILGYMVSSTSELMKKNMNTDEHNDYET
NRCDEETEDLPILYNNIKYNNKINCYTPHAILNLITLNDENTLNNKNSDNK
RLNVDKFKFLVDIYDITTYITKSLDKLKEITNLTSTNHSNKKNNKILAYK
KYALLKILKRGFNICYIADTYVYNNKRSKSDLSSEINLILNLDLKLQKRNK
LEPHPTQNTTQIKINIKNKLILYHQNKKVKKIIFKTKITKNLQDQRYHSE
NKRKITKDSLLAINHKTNIKIEQKISITSNELSKLRFSL"
gene complement(20528. .21454)
/gene="MAL3P5.5"
CDS complement(20528. .21454)
/gene="MAL3P5.5"
/note="predicted using hexExon; MAL3P5.5 (PFC0595C),
Serine/threonine protein phosphatase (PP2), len: 309 aa;
Similarity to serine/threonine protein phosphatases.
M.domestica serine/threonine protein phosphatase
(Ts:Q42912) BLAST Score: 1005, sum P(1) = 6.9e-107; 60%
identity in 301 aa overlap."
/codon_start=1
/protein_id="CAB38970.1"
/db_xref="GI:4493934"
/db_xref="SPTRMBL:O97259"
/translation="MAKGEERKIBQLRMPPKLLDESDDLQVQKVELLVEENVO
SIKPPVILGDIHGQFDLLEPFDVGDDIMNDYIFLGDYDGRGNSVETPILLLK
LLFPKNTILRGNHESRQITTYGVQDECFKYGANAWKICTDI:FDYTLAALVDNQ
IYVGHGSLSPKILIDQLALINRVQIEPHGAPGDLMKSDPDEVDVWNPAGQWGLP
GPVITKKNFINLLELILARHOLAMEGRTYMFEDSTIITVWSPNTCTRCGNAALMR
IDSEYNMGLMFKDTPDSRNSIKNATIPFL"
gene 25252. .26157
/gene="MAL3P5.6"
CDS join(25252. .25296,25453. .26157)
/gene="MAL3P5.6"
/note="predicted using hexExon; MAL3P5.6 (PFC0600W),
Hypothetical protein, len: 250 aa"
/codon_start=1
/protein_id="CAB38972.1"
/db_xref="GI:4493936"
/db_xref="SPTRMBL:O97261"
/translation="MKYILKNWYIYNNYIKNLEEKYNFLKLNNSHMGASQNNVN
NWTNMLEEPEKINNNYNNNNNNNNYNDYDNLNIVSQVQNNKRLQDFYNNKN
SPGYIKLKTORFDDIDILNLEKRLAYEROWLLKINQEEKNKIGINGPGESE
NSSSSNNNNYLLYKINLAKTWNKSKRSKKRRKINSIKDKYIKRCKACTINP
NGKFIEDYITTCYQNDVSIRSPNNAD"
gene 25747. .26290
/gene="MAL3P5.7"
CDS 25747. .26290
/gene="MAL3P5.7"
/note="predicted using hexExon; MAL3P5.7 (PFC0605C),
Hypothetical protein, len: 248 aa"
/codon_start=1
/protein_id="CAB41709.1"
/db_xref="GI:4725991"

```

/db\_xref="SPTRMBL:O9Y011"

```

/translation="MGHSGSLIKPQKNWYKQAKQKVKVYDEHRIIEKDKREIR
KADPEFSTISLAKNKNKEDSDNNIHFIDEMGEKKTITNTCKDNLNLFIDEEKL
ADKAREBFLIKGHTYITDKNFNTQNNISIDWNKQIISIDPMKNKCFDWFNLK
NKEATDQNGAFNPIQDNIISEENHETINSDLSICTNWNITIDAKKQKQMY
HKAIIKIFQEKREKRRKQKKAAPK"

```

complement(29992. .33573)

/gene="MAL3P5.8"

complement(29992. .33573)

/gene="MAL3P5.8"

/note="predicted using hexExon; MAL3P5.8 (PFC0610C),

Hypothetical protein, len: 1182 aa"

/codon\_start=1

/protein\_id="CAB38971.1"

/db\_xref="GI:4493935"

/db\_xref="SPTRMBL:O97260"

```

/translation="MAKVKKKETEAGTQPVVAKBQTHAKENNESNIAVTENVIS
KNQDEIISKQDEIATSKNDQEIASINQDEIATSKNDQENVALSKSEERQNAKSE
DNELAQIKPEKTSNENHNNNSSTSTLSSFFPEYENSVQHPFSGTTEHSEME
DSNNVEIENAIINVDNRSSTSYSKQKLEISVTCVTCQYDNLINSDHFAHL

```

```

Query Match 6.18; Score 185.6; DB 33; Length 86829;
Best Local Similarity 45.6%; Pred. No. 6.2e-12;
Matches 1207; Conservative 1; Mismatches 1396; Indels 45; Gaps 14;

```

```

Qy 138 aatataataacatogtagaataaattttatccaattgaagtattcaacatttta 197
Db 38731 AATTTCATAAAGTACACAAATAATATATATATAGTATAAGAAAGATTAAAGACATTA 38790
Qy 198 atattgtagatgtaatttaaatgaagataaataattcttgacatgtattttct 257
Db 38791 TTAATTATATATACATTTTTTATTAAATATATATATAATAATATATATATACAA 38850
Qy 258 ctaaatgtttgtgctttgtgtgtaggtgtatgtagtgcattctttaaacaata 317
Db 38851 AT-ATAAATCATGCTTTAAATAATAATATATATTTTATTATATATATAATAGTA 38908
Qy 318 tcacettttgagttgtatgatgataagatgaataaagaaatgggtgatcttcc 377
Db 38909 TTAATTATATGTTATATACATTTTATTAACTCTTTCTAACTTAATATGATGACATA 38968
Qy 378 tttagacttgtagtaagacaaacttcaaaagattgtatgtatcatatataata 437
Db 38969 TTTTATATATAATTTTACACATTTATTAATAAACATATATATATATATATATTA 39028
Qy 438 tatcttcaaaatttataataaaattgtgttaataaattacagtataatttttt 497
Db 39029 TTTTAAATAAAATATAATATTTTATAAATAATAAAATAATATATATATATATAT 39088
Qy 498 atctctaattttattgtgcgaatttttagatttatattacaataaaaaattgta 557
Db 39089 ATATATATATATTAATATATATAAATTAATATATATATATATATATATATATAT 39145
Qy 558 caacttacaagccatacaataaattataataattcatataaanaattatataaa 617
Db 39146 TATATTAATAAATAAATAAACAATAAATAAATATATATTTTAAATTAATAAATA 39205
Qy 618 taacagtataaataaactatttagaattattctattcttaagatacaactgaatt 677
Db 39206 AAAATATATATTTTATAAATAAATAAATAATATATATATATTTTAAATAAATA 39265
Qy 678 gtataaataaagttgattgtatttagaagtgagtgattattctgtataacataatcc 737
Db 39266 AATTATATATATATTTTATAAATAAATAAATAATATATATATATTTTAAATAA 39325
Qy 738 taacatttttataaacttctgttttgaaagttccaaagaagaaaggaagaaatt 797
Db 39326 AAATAAATAAATAATATATTTTATAAATAAATAAATAATATATATATATTTA 39385
Qy 798 gagagtaagtctgatttatataatacaataagagtgattttctcttttataat 857
Db 39386 AAATAAATAAATAATATATATTTTATTAATAAATAAATAAATAATATATATAT 39445

```



```
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AC47815.1"
/db_xref="GI:1166533"
/translation="MSTRKSNPPHLDVSPWFLGGAIGAMTIVSGMWHQYDLSL
VLGNITILILYQWNRDVRSGTQGLHTAVTIGLWQWHLFLSEVLFYSPFWAF
PSSLSIAELGASWPMGIGISFPFOIPLNTALLASGVVYWAHSLMNNESQT
TQGLFPTVLIGIITPILQAYTIAAPPTIADSGTSTFPAATGSGHVLGTTFLV
CLLRHLNWFKNHFGPEAANTWHPDVKFLITITWGG"
tRNA 5543..5607
/gene="mt:ND6"
```

Query Match 5.48; Score 193.8; DB 58; Length 19517;  
Best Local Similarity 45.38; Pred. No. 1.3e-12;  
Matches 1291; Conservative 0; Mismatches 1490; Indels 59; Gaps 14;

```
Qy 87 tctatattgtcttccctagctggccacgcgaaattttctgtgtacaaataata 146
Db 18579 TTATATTTCATATAAAATTTATTATATAAAATTTTGTGTTATTTTAAAAACATG 18520

Qy 147 aatacatcgtagaataaattttatccaattgaagcttaaacatttttaatttgta 206
Db 18519 ATTTATATATATAAATTTTATATAAAATATACATTATAGAAATTTTAAAA----A 18465

Qy 207 gatgttaattaaatgaagataaatacatattctggacatgattttcattotta-atg 265
Db 18464 AATTATATTAATTAATTAAATAATTAATTTTCTATATATATATATATATATAA 18405

Qy 266 ttgtgctttggtgataggtgtattgatgcacagctgtctttaaatacacaatcatt 325
Db 18404 ATATTCAATATAATAAATTTATATAATATATAAATAATTAATTAATCAATATATATT 18345

Qy 326 ---tgagttgatgatgaagtcgaacataaagcaaatatgggtgatcttcatttg 382
Db 18344 ATATAAATAAATTAATAATAAATAAATAAGAAATATAAATTTTATAAATAATATCTA 18285

Qy 383 aacttgcataagtcacacaactttaacaagcttgattggtgcacataataataatct 442
Db 18284 CATTTTAAATTTAAATTTTTTATTAAATATTAGATATATAAATAATATAAATA 18225

Qy 443 tcaattttataataaaaattggtttaaataattacagttatatttttttatctc 502
Db 18224 TTTATATATATAAATATCTATTAAATTTAATATAGTATATAGTTTTTTTAAAAAAA 18165

Qy 503 taattttattgtgcgaatttttagttgatatttaacataaaaaaattggtacacat 562
Db 18164 AAATTATTTTTTAAAAAATTTTTTAAAAAATGAAATAAATAAATATATTTCAT 18105

Qy 563 ttacagcccatatacaaaataattataaatattcattaaaaaata-----tatt 613
Db 18104 TATAAATTTTATTATATAAATTTTGTGTTATTTTAAAAAATGATGTTTATATA 18045

Qy 614 taatataggataaataataa-ctatttagaattattctacttaagaataacataggt 672
Db 18044 TAAATATTTTTTAAAAAATAACATTATAGCAATTTTAAAAAATTTATATTAATTA 17985

Qy 673 taattgtataattaaatgaaggtagtttatgtgaagatgaggtatattgtgaaacata 732
Db 17984 TTTAAATAATTAATTTTCTATATATATATATATATATATAAATTCATATAATAT 17925

Qy 733 atcaatacaacctttttatacaactttctgttggaagtcacaaagaagaatgaagga 792
Db 17924 AAATTATAAATATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17865

Qy 793 aatttgagagtaagttcattttatataacataatgaaggtgatgtttctctctttt 852
Db 17864 TATAAATAAATATAGAAATATAAATTTTATAAATATATCTATCAATTTTAAATTTAA 17805

Qy 853 aatatttttatacaaaattttaataaaataaataaagattgaagaaataaataga 912
Db 17804 AATTTTTATTAATTAATAGATATATAAATAATTAATTAATTAATTAATTAATTA 17745

Qy 913 aagctgttttactaatgacatattgattttgtgcacatacttaataataagat---a 969
```

```
Db 1774 TATCTATTAATTATATATAGTATATAGTTTTTTTTTAAAAAAAATATTTTTTTAA 17685
Qy 970 aattaatgttggtacattgatacaagaacaacatagatttgcacattctattgttaa 1029
Db 17684 AAAATTTTTTTTTTAAAAATGAAAAATAAATAATATATTTTCATATAAAATTTATTA 17625
Qy 1030 aagctgctgcgtttacataaaataaggtacatgtttacatgcacatacaactatcggt 1089
Db 17624 TAAAAATTTTTGTATTTTTTTAAAAAATGATTTTTTATATATAAATTTTTTATAA 17565
Qy 1090 tattctatacaatcgcctaatttttacagatagaatgaatgaatttttaataagaag 1149
Db 17564 AATATACATTATAGAATTTTTTAAAAAATTTATATAAATTTATTAATTAATTTAAT 17505
Qy 1150 ggtcaaatgtttattgatctaacacgtaggatttaatttacttattt-----tctaaa 1204
Db 17504 TTCTATATATATATATATATAAATAATCAATATATATAAATTTATATAATATATA 17445
Qy 1205 gaataaggaataataaattgaatcttaatacaaaactttctgatacttttatcata 1264
Db 17444 TAAATTAATTAATTTATAAAAAAAATAAATAAATAAATTTTATTAATTAATTA 17385
Qy 1265 ttttactataaatttaattgtgagatgaacaatataaaactgaagaacccaana 1324
Db 17384 TATTAGTATAAATAAATTTTATTAAATTTAGTATTAATTTATTATTTATTAATA 17325
Qy 1325 gttagttatgttgatcatatacaacagtttaaaattgaataaattttttctogtaa 1384
Db 17324 AATATATAAAG-----AATAGATTTTATTATTAATTAATTAATTTATTTATA 17270
Qy 1385 ttaattccatctgggtttttttttctagtaagcaataatacaaaataaataca 1444
Db 17269 TTAATAATTTAAATTTATTTCATTTAATAATATATATATATATATAAATTTTAA 17210
Qy 1445 ttaattccatcaatacccgccgtgcctccctccctcaacttaaaacaa-actaaca 1502
Db 17209 TTAATTAATTAATAATTTTATTATAAATAAATTTTATAAATAAGTTTATAGT 17150
Qy 1503 ccagcagaacaacgcactttaataagccactttctagcagatgtcttgcaactaaga 1562
Db 17149 ATAAATTTATAAATAATTTTTTTTAAAAAATAATTTTTTAAAGTTTAAATATA 17090
Qy 1563 aagtaagctaacgcgcaactcaatccatcagagcccaacagatagaaggtgttgat 1622
Db 17089 TAAATTTATGAATAGGGGGAATAAATTTATTTTCATTTTATATATATATATATA 17030
Qy 1623 ggtttgcacaaagttgttaaacccggcccaactccctttttcttcaatccccc 1682
Db 17029 TATATACATTAATTAATTCAGATTTAGTATTAAATAAATTTTATTTATATATAT 16972
Qy 1683 actccacacccctcaattttcttctatgtgtcttattataagttcttataatcagaa 1742
Db 16973 --TATATAATTAATTGAAGAAATAAATATATGATATATAAATAATGAATGA 16916
Qy 1743 tcaagataagctctcgcaacaacaaacccatggctctcgagcaagatgtgactatga 1802
Db 16915 TTTTATAAAAAATCATTTAAATTTTATATATATAAATAATTTTATTAATTTATA 16856
Qy 1803 gactcgtgaatttgatcattatacagcaaaacagtttaacaaagctgttcagat 1862
Db 16855 AATATAATTTTATATAAATAATTTTATATAAATAAGTTTATTAAGTATAATTTAATA 16796
Qy 1863 aaacactgaatcgtcctag-----ttgtttgtgttcaatcagttccacaggaact 1917
Db 16795 AATCATTTTTTTTAAAAAAAATAATTTTAAAGTTTAAATATACATAAATTTATGA 16736
Qy 1918 atgagcctctctaaagaacacacacatactcaacacccctgcatacagtgatgt 1977
Db 16735 ATAGGGGGAATAAATTTATTTTCATTTTTTATATATATATATATATATATAA 16676
Qy 1978 tcaattttgcagataaagttcttttttaactctcttttactcatttttaaacagat 2037
```

Db 1224 AAAATAGTTTATTAGTATAAATTAATATCATTTTTTTTAAAAAATAATTTTT 1165  
 QY 2569 ATTTTATTTTATTTAGAAATTTTATATGACCTCACTCACTAGTAAAGTTGTTATA 2628  
 Db 1164 AAGTTTAAATATACAAATTTATGATAGAGGGGGAATAAATTTATTTTCATTTTTTA 1105  
 QY 2629 TAGAGTAGTGTATAATATATAGTAGTATTTTGTCTATAAACAATAAATGGG 2688  
 Db 1104 TATATATATATATATATAAATTAATTTATTCAGATTTAGCTATAAAATAAATTTAT 1045  
 QY 2689 TCTTGGTGGCTGAGCTGAGCTCAATGAGTCAATGAGCTAATCTCATTTTAAAC 2748  
 Db 1044 ATTAATATTATATATATTTTGAATTTGAAATTTAAATTTATATATATATATATAAATA 985  
 QY 2749 AGGCTGATTTTATTTTATTTATTCAGCTGAAATTTTTCGGTGAATCTCTCGAGCTCA 2808  
 Db 984 TAAATGAAATTTTTTAAAAATATTTTATTTTATTTTATATAAATAAATTTCTTATTA 9205  
 QY 2809 GATTAATCAACCAAGGCTCAATGTAGTCACTGAAGAAATGAATCTAGTCACTCA 2868  
 Db 924 ATTTTATTTAAATTAATTTATATATATAAATTTATATATAAAATAGTTTATATAGTATA 865  
 QY 2869 ATTAATATCTCATCTCTCTTTGGTGA 2896  
 Db 864 ATTTAATAAATAATTTTTTTTTTAAAAA 837

RESULT 3  
 NC037541.1  
 LOCUS Drosophila melanogaster complete mitochondrial genome. 04-APR-2000  
 ACCESSION U37541  
 VERSION U37541.1 GI:1166529  
 KEYWORDS  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 12511 to 12682)  
 AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.  
 TITLE Drosophila mitochondrial DNA: a novel gene order  
 JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)  
 MEDLINE 83090428  
 REFERENCE 2 (bases 5269 to 5695)  
 AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.  
 TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes  
 JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)  
 MEDLINE 83220794  
 REFERENCE 3 (bases 404 to 5272)  
 AUTHORS de Bruijn,M.H.  
 TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code  
 JOURNAL Nature 304 (5923), 234-241 (1983)  
 MEDLINE 83245048  
 REFERENCE 4 (bases 804 to 1778)  
 AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.  
 TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species  
 JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)  
 MEDLINE 88174373  
 REFERENCE 5 (bases 5268 to 13619)  
 AUTHORS Garesse,R.  
 TITLE Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations  
 JOURNAL Genetics 118 (4), 649-663 (1988)  
 MEDLINE 88212147  
 REFERENCE 6 (bases 441 to 2967)  
 AUTHORS Satta,Y. and Takahata,N.

```

c 25 134.8 4.4 130117 39 AC004907 AC004907 Homo sapi
c 26 133.8 4.4 158116 60 AC006281 AC006281 Plasmid
c 27 133.2 4.4 2426 8 SDH49822 U49822 Saccharomy
c 28 133 4.4 170807 39 AC005089 AC005089 Homo sapi
c 29 132.2 4.3 106550 39 AC007708 AC007708 Homo sapi
c 30 132.2 4.3 176552 39 AC004617 AC004617 Homo sapi
c 31 132.2 4.3 203519 40 CNS01RHO AL162191 Homo sapi
c 32 131.8 4.3 321003 31 PPMAL4P3 AL035476 Plasmid
c 33 131.2 4.3 80518 31 PPMAL13PA AL109815 Plasmid
c 34 130.6 4.3 170125 41 AC007465 AC007465 Homo sapi
c 35 130 4.3 321003 31 PPMAL4P3 AL035476 Plasmid
c 36 129.6 4.3 282806 60 AC006279 AC006279 Plasmid
c 37 129.4 4.2 158116 60 AC006281 AC006281 Plasmid
c 38 129.2 4.2 224448 31 PPMAL4P4 AL035477 Plasmid
c 39 129 4.2 318221 31 PPMAL13P3 AL049184 Plasmid
c 40 128.8 4.2 106550 39 AC007708 AC007708 Homo sapi
c 41 128.8 4.2 176552 39 AC004617 AC004617 Homo sapi
c 42 128.6 4.2 12029 34 AE001372 AE001372 Plasmid
c 43 127.6 4.2 75076 39 AC004946 AC004946 Homo sapi
c 44 127.6 4.2 204652 31 PPMAL13PE AL049183 Plasmid
c 45 127.2 4.2 12029 34 AE001373 AE001373 Plasmid

```

## ALIGNMENTS

```

RESULT 1
S79308
LOCUS S79308 913 bp mRNA PLN 30-NOV-1995
DEFINITION Rac13>21.8 kda GTP-binding protein [Gossypium hirsutum-cotton
plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 913 nt].
ACCESSION S79308
VERSION S79308.1 GI:1087110
KEYWORDS
SOURCE upland cotton boll fibers cv. Acala SJ-2.
ORGANISM Gossypium hirsutum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurasids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 913)
AUTHORS Delmer,D.P., Pear,J.R., Andrawis,A. and Stalker,D.M.
TITLE Genes encoding small GTP-binding proteins analogous to mammalian
rac are preferentially expressed in developing cotton fibers
JOURNAL Mol. Gen. Genet. 248 (1), 43-51 (1995)
MEDLINE 95379748
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 170155] from the original journal article.
This sequence comes from Fig. 1A.

```

```

FEATURES
source Location/Qualifiers
1..913
/organism="Gossypium hirsutum"
/db_xref="taxon:3635"
gene 12..602
/gene="Rac13"
/note="21.8 kda GTP-binding protein"
12..602
/gene="Rac13"
/note="21.8 kda GTP-binding protein; pea RhoL protein
homolog/mammalian rac protein homolog; ; This sequence
comes from Fig. 1A"
/codon_start=1
/protein_id="AAB35093.1"
/db_xref="GI:1087111"
/translation="MSTARFIRCVTVGDAVGKTKQLSYISNTFFPDIVYTVDFNFS
ANVVVDGTVNLGLWDTAGQEDYNRLRLSTYRGADVFLAFSLISASTYENIKWKP
ELRHYAHNVFVVLGKTLADDDKQFLIDHGPATPISQGEELAKMIGAVTYIECS
KIQQNVKAVDAALIKVALRPPKPKKPKRRCARL"
BASE COUNT 307 a 169 c 172 g 265 t
ORIGIN

```

Query Match 10.04; Score 304.8; DB 8; Length 913;  
 Best Local Similarity 97.84; Pred. No. 8.5e-24;

```

Matches 309; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1806 cctcgtgatttgatcattattacagtcacaaacagtcacaaagctgctgcagataa 1865
Db 597 CTTTGAATATGGCATCATATACAGTCACAAACAGTTACACAAAGCTGTTGCAGATAA 656
Qy 1866 cactgactgctgctatgctgtttgtttgtttacatattgttccagtcgaactatgaaga 1925
Db 657 CACTGAATCTGCTATAGTGTGTTTGTGTTTACATATGTTCCAGTGAACTATGAAGA 716
Qy 1926 tctctaaagaaacacaaactacatcaacacctgacatgacatgcttcaatttt 1985
Db 717 TCTTAAAGAAACCCAACTATCATACCACTGCATGATGATTCATTTT 776
Qy 1986 cgcagataagttcttttttaactcttttactcttatttaacgaattcatgaa 2045
Db 777 CGCAGTATAGTTCCTTTTAACTCCTTCTTTTACTCTATTTATGCGAATCTATGGA 836
Qy 2046 taattgtccctacaaactgctacaaactgttaattataataattcattctatttt 2105
Db 837 TAAATGTCCTCAACACATGTCATTAAGTGTTAATATTAATTCATCTCTATTTT 896
Qy 2106 actaagatattagtaa 2121
Db 897 ACTAAAAA 912

```

```

RESULT 2
DM011584/c
LOCUS DM011584 4601 bp DNA INV 23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION U11584
VERSION U11584.1 GI:508826
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.
SOURCE fruit fly.
ORGANISM Mitochondrion Drosophila melanogaster
Bukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 4601)
AUTHORS Lewis,D.L., Parr,C.L., Parquhar,A.L. and Kaguni,L.S.
TITLE Sequence, Organization and Evolution of the A+T Region of
Drosophila melanogaster Mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11, 523-538 (1994)
MEDLINE 94285822
REFERENCE 2 (bases 1 to 4601)
AUTHORS Kaguni,L.S.
TITLE Direct Substitution
JOURNAL Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D., Dept. of
Biochemistry, Michigan State University, East Lansing, MI,
48824-1318, USA

```

```

FEATURES
source Location/Qualifiers
1..4601
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/strain="Oregon-R"
/db_xref="taxon:7227"
/dev_stage="embryo"
gene 1..4601
/gene="mt.ori"
/note="mitochondrial origin"
/allele=""
/db_xref="FlyBase:FBgn0013687"
repeat_unit 650..1022
/gene="mt.ori"
/note="repeat 1-A"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
repeat_unit 1023..1360
/gene="mt.ori"
/note="repeat 1-B1"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem

```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```
FEATURES
  source
    Location/Qualifiers
      1..1101
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="RPCI-98"
      /clone="BACR05K11"
      /note="end : TET3"
BASE COUNT      631 a      7 c      28 g      289 t      146 others
ORIGIN
```

Query Match 2.04; Score 111.8; DB 122; Length 1101;  
Best Local Similarity 41.84; Pred. No. 8.6e-08;  
Matches 270; Conservative 63; Mismatches 312; Indels 1; Gaps 1;

```
Qy 4759 ttttaaatgttgatctaatgttaacatgaactgttgctgattatgttattgtgtatt 4818
    ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1065 TWTTTNTWTCCTCTTTTTTTTTTTTTTTTTTTTTTMMWTTTMMGAATATTMCACCTTTT 1006

Qy 4819 ttacttttaagtattgtcatgtattgtttaaacattgctgacattatctctct 4878
    : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 CATMCTTTCATATTTTTCMTMHTTTTTTMMWTTTMMAMMTTATMCTTACHATTTTAA 946

Qy 4879 actatttaataaataagcactgtttgttttaaacctttttacaagtttaagacatgtataa 4938
    || : ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 ACMTTMMWTTTMMCMWMTTTCMCATTTTTTAACTTAAAAAATATNTAAATSAATAT 886

Qy 4939 atatatgaacataataattacaggttttagtctaatgttagctatcttagtattgtatga 4998
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 885 TAAATTTTTTAAATTAANAATAATTAAAAAAAWMTTWTWTWTWTWTWTWTWTWTAT 826

Qy 4999 tgaatttaataacatttaacaaattccactttaaaattttaaataataacaaataat 5058
    | : || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 825 TWTAN 766

Qy 5059 tattgttaataataacattaaatgaacaaaaaataaataaataaataaataaataa 5118
    ||| : ||| ||| : : ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 765 TATWTAAAAAATTTAAWMTWTATWTWTAAAAAWTATTAATWTAAAAAATTTWTAATWTAN 706

Qy 5119 taattgttataatattgttaataataatgtacacattcttaactgaatagggtctaac 5178
    || : || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 705 AAAATTTTTTAAATTTTATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTAT 647

Qy 5179 ctataatccctaaattttcagttttaaatattttataccatattattagaactctt 5238
    : || ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 AWAATAAAAAAATWTWTWTAAAAWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTAT 587

Qy 5239 tttaaatatattaaatttttaattataccaatatttaaacattattattattatttaacta 5298
    | | : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 TAAATTTTWTWTAAAAATTAATAATTAWAAAAAATAAAAAAATAAAAAAATAAAAAAATA 527

Qy 5299 aaactcaaaattttatttaacattataaaattcctaatattcttatctataattaaac 5358
    || : || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 AAWTAAATTAATATTTTWTIN 467

Qy 5359 tctaatattcctaaatttaattaaattcttaattattcttaattgt 5404
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 TTTTWT 421
```



please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Gotozawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://dceac.med.buffalo.edu/drosophila/bac.htm>.

| FEATURES              | Location/Qualifiers   |
|-----------------------|-----------------------|
| 1. <b>NAME</b>        | 1. <b>NAME</b>        |
| 2. <b>ADDRESS</b>     | 2. <b>ADDRESS</b>     |
| 3. <b>PHONE</b>       | 3. <b>PHONE</b>       |
| 4. <b>FAX</b>         | 4. <b>FAX</b>         |
| 5. <b>E-MAIL</b>      | 5. <b>E-MAIL</b>      |
| 6. <b>WEBSITE</b>     | 6. <b>WEBSITE</b>     |
| 7. <b>DESCRIPTION</b> | 7. <b>DESCRIPTION</b> |
| 8. <b>CONTACT</b>     | 8. <b>CONTACT</b>     |
| 9. <b>NOTES</b>       | 9. <b>NOTES</b>       |
| 10. <b>STATUS</b>     | 10. <b>STATUS</b>     |
| 11. <b>REMARKS</b>    | 11. <b>REMARKS</b>    |
| 12. <b>DATE</b>       | 12. <b>DATE</b>       |
| 13. <b>TIME</b>       | 13. <b>TIME</b>       |
| 14. <b>LOCATION</b>   | 14. <b>LOCATION</b>   |
| 15. <b>QUALIFIERS</b> | 15. <b>QUALIFIERS</b> |

```

NAME      1. .1101
source
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29P01"
/notes="end : TET3"
BASE COUNT      366 a      66 c      104 g      351 t      214 others
ORIGIN

```

Query Match 2.1%; Score 115.6; DB 122; Length 1101;  
Best Local Similarity 40.6%; Pred. No. 2.3e-08;  
Matches 231; Conservative 81; Mismatches 255; Indels 2; Gaps 1;

[illegible]

Qy 4763 aaatgttgatctaagttaacatcacttggcttgatttatgtagttatgtatatttac 4822  
| : | : | : | :: || | : :|| : : | : || : ||

Qy 4823 tttaatgatattgcgatgtattgttaattaacattgcttgatcattatactcttctacta 4882  
: || :||| | | :| | : | ||||: : | | ||

Qy 4883 ttaattataaatggcactggtttgtttaaacgtttttacaagttaagacatgtataaatat 4942  
: | | | | : | : | | : | | | | : | | | |

Qy 4943 atgacaatataattacaggttttagttcaatgtagctatcttagtatgttattgatgat 5002

Qy 5003 ctttaattacatttaaacaaattccacttaaaattttaataaataataacaaataattatt 5062

Dd 671 WTAATTAATAAATWATTTAAATWTAAATATAAATTAAADRTTAAAWTTTTTTT 612

Gy 5063 gtaataataacattaaatgcaacaaaaaatgaataaataaaaataaaatagcaataat 5122

Db 611 WNCWTTNAAATATATAAAATWAAAAATTAAATTWTTTTATATWWATAAAAAAMATWWNTT 552

Ov 5123 ttttataatattgtaatataaatatgtaccattattcttaactgaatacgggtctaacctat 5182

Db 551 ATTTATAAAWAAATTATTAAWATTGATTTTATTATTWTTTTTAAWTTTTATTATA 492

Db 491 TWATTATTAAATWTTAWTATATTTTTTTTATATCTTTCMAAWTATTTMTCCCCYTTTW 432

Db 431 TTATGTTTTTTTTTTHCTCTTNCNTT 403

```

RESULT 10
B11102/c
LOCUS      B11102      1187 bp    DNA                GSS            14-MAY-1997.
DEFINITION F19C22-T7 IGF Arabidopsis thaliana genomic clone F19C22,
            genomic survey sequence.
ACCESSION  B11102

```

```

VERSION      Bill102.1  GI:2092386
KEYWORDS     GSS.
SOURCE       thale cress.
ORGANISM     Arabidopsis thaliana
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
              Brassicaceae; Arabidopsids.
REFERENCE    1 (bases 1 to 1187)
AUTHORS      Peng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
              Ecker,J.
TITLE        Arab. Ec Sequences at ATGC
JOURNAL      Unpublished (1997)
COMMENT      On Sep 10, 1998 this sequence version replaced gi:3556525.

```

Other\_GSSs: F19C22-Sp5  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
19104  
Tel: 215-898-9384  
Fax: 215-898-8780  
Email: jacker@atgenome.blo.upenn.edu  
Seq primer: 77  
Class: BAC ends  
High quality sequence start: 72  
High quality sequence stop: 353

| FEATURES | Location/Qualifiers |
|----------|---------------------|
|----------|---------------------|

```

source      1. .1187
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="E19C22"
            /clone_lib="IGF"
            /sex="hermaphrodite"
            /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
            Produced by Thomas Altmann"

```

BASE COUNT 385 a 51 c 60 g 594 t 97 others

Query Match 2.1%; Score 115.2; DB 120; Length 1187;  
Best Local Similarity 47.0%; Pred. No. 2.7e-08;  
Matches 362: Conservative 0; Mismatches 402; Indels 6; Gaps 2;

Qy 3208 ttatattacggaatgtaattatatttttaaaataaaattatgttatttagattottaat 3267

Qy 3268 attttgagcattccatactataatttcgtaacataatattaaaatatagtaataaaag 3327  
 || | || ||| || | || | ||| || ||||| || |||

Qy 3328 tgtaattaactttaattacaagcataatattaaatttgaatcaattaattttatttc 3387

Qy 3388 tattattttaattaatttagtctatttttcaaaataaaatttaaatctaaataaaaata 3447

Qy 3448 attttccctaagtgtgaaacaactcatgttatacttcaaaattataagtattatattta 3507

Db 945 TATTTTATTAAAAAANAANNAATATTTTTATTAATATTAAATATAAATAAAA 886

Qy 3508 ccttgatgattttatttagtatattaattctgattataattatggtgggatacaatcg 3567

Db 885 AAAATTTTTTTTAAATATAATAATAAAAAATATAAANAATAANAATAAAAAATATAT 826

Db 825 TTTAAAAAATTAAATATTTAAAAATAATTAAAAAATTATTTAAAAATTAATATTTATNANA 766



BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genet cs ([info@regen.com](mailto:info@regen.com)). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: Sp6  
Class: BAC ends

FEATURES  
source Location/Qualifiers  
1..718  
/organism="Homo sapiens"  
/db\_xref="GDB:757344"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-1928"  
/clone\_lib="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC11 Human Male BAC Library"  
BASE COUNT 310 a 39 c 35 g 334 t  
ORIGIN

Query Match 2.14; Score 117.2; DB 102; Length 718;  
Best Local Similarity 48.84; Pred. No. 1.4e-08;  
Matches 317; Conservative 0; Mismatches 333; Indels 0; Gaps 0;

Qy 474 ttgtatgaatttaatttaattgtgtatcaattgttaactcaattgctgtttat 4803  
|||||  
Db 17 TTGTGCTTTTCTGTGTGAAGTTTTTGACTACGATGAATCTCTTACTAGCATAA 76  
  
Qy 4804 gttatgtttgtttatttcttaattgtatgttcattgttttttaacattgttga 4863  
|||||  
Db 77 GTATTTCAGCAATCTCTTCTCTGTGATGTCATTTGTATATTCATGTTCTTATGT 136  
  
Qy 4864 tctattactctctctacttaatttaataatgacattgtttgttttaacattttacag 4923  
|||||  
Db 137 ATATTATATATATATATATATATATATATATATATATATATATATATATATAT 196  
  
Qy 4924 ttaagacattgataaattgacacataataacacagttttgttcaattgttgcattc 4983  
|||||  
Db 197 ATATTATATATATATATATATATATATATATATATATATATATATATATATAT 256  
  
Qy 4984 ttgtatgtttgtatgtcttaattcaatttaacaaattccatttaatttaattat 5043  
|||||  
Db 257 TAATATATGTTAATATATATATATATATATATATATATATATATATATATATAT 316  
  
Qy 5044 ataatacaaatatttattgtataataacatttaattgacaaataattgaataata 5103  
|||||  
Db 317 ATAATATATATATATATATATATATATATATATATATATATATATATATATAT 376  
  
Qy 5104 aaataaaatgacaaatttatttataattgttaataataattgtacattatttaact 5163  
|||||  
Db 377 ATATTATATATATATATATATATATATATATATATATATATATATATATATAT 436  
  
Qy 5164 gaattaggttgaacttaacttaacttaatttttgaatttttttaacttaact 5223  
|||||  
Db 437 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 496  
  
Qy 5224 attattgaactcttttttaatttaataaaatttttaacttaacttaacttaactat 5283  
|||||  
Db 497 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 556  
  
Qy 5284 taattctatttaacttaacttaacttttttaacttaacttaacttaacttaacttct 5343  
|||||  
Db 557 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 616  
  
Qy 5344 tatcttaatttaaaactcttaacttaacttaatttttaatttaatttttaacta 5393  
|||||  
Db 617 ATTATATATATATATATATATATATATATATATATATATATATATATATATAT 666

RESULT 7  
B11102  
LOCUS B11102 1187 bp DNA GSS 14-MAY-1997  
DEFINITION F19C22-77 1GF Arabidopsis thaliana genomic clone F19C22,

genomic survey sequence.  
ACCESSION B11102  
  
VERSION B11102.1 GI:2092386  
KEYWORDS GSS.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;  
Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 1187)  
AUTHORS Peng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and  
Ecker, J.  
TITLE BAC End Sequences at ATGC  
JOURNAL Unpublished (1997)  
COMMENT On Sep 10, 1998 this sequence version replaced gi:3556525.  
Other\_GSSs: F19C22-Sp6  
Contact: Ecker J.  
Arabidopsis Thaliana Genome Center  
University of Pennsylvania  
Dept. of Biology, University of Pennsylvania, Philadelphia, PA  
19104  
Tel: 215-898-9384  
Fax: 215-898-8780  
Email: [jecker@atgenome.bio.upenn.edu](mailto:jecker@atgenome.bio.upenn.edu)  
Seq primer: T7  
Class: BAC ends  
High quality sequence start: 72  
High quality sequence stop: 353.  
  
FEATURES  
source Location/Qualifiers  
1..1187  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="F19C22"  
/clone\_lib="1GF"  
/sex="hermaphrodite"  
/note="Vector: BeloSACII; Site\_1: EcoRI; Site\_2: EcoRI;  
Produced by Thomas Altmann"  
BASE COUNT 385 a 51 c 60 g 594 t 97 others  
ORIGIN

Query Match 2.14; Score 117; DB 120; Length 1187;  
Best Local Similarity 43.64; Pred. No. 1.4e-08;  
Matches 365; Conservative 0; Mismatches 470; Indels 3; Gaps 2;  
  
Qy 3129 agaatattgttaaaatgaatttaaaatgaattgctggtgacacacacaaataaact 3188  
|||||  
Db 328 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 387  
  
Qy 3189 atgttgttgggtgaattttatattacggaattgataattatttttaaaataaata 3248  
|||||  
Db 388 ATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 447  
  
Qy 3249 tttatttagatttcttaattttttgagacattcactataattctgaacataatt 3308  
|||||  
Db 448 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 507  
  
Qy 3309 aaatatagtaataaaagtgaatttaacttaatttaacagcataatttaattttga 3368  
|||||  
Db 508 AAGTTTATTAATTTTAAATTAATTAATTAATTTTAAATTAATTAATTAATTAAT 567  
  
Qy 3369 atcatttaatttttttttttttttttttttttttttttttttttttttttttt 3428  
|||||  
Db 568 AAGTTTATTAATTTTAAATTAATTTTAAATTAATTTTAAATTAATTTTAAATTA 627  
  
Qy 3429 ttaatttaataaaataatttttttttttttttttttttttttttttttttttttt 3488  
|||||  
Db 628 ATTTTATTAATTTTAAATTAATTTTAAATTAATTTTAAATTAATTTTAAATTA 686  
  
Qy 3489 attatagattattattacattgatttttttttttttttttttttttttttttttt 3548  
|||||

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES             Location/Qualifiers
     source            1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR08K10"
                        /note="end : TET3"
BASE COUNT            201 a      64 c      131 g      202 t      503 others
ORIGIN

```

Query Match 2.3%; Score 126.2; DB 122; Length 1101;  
Best Local Similarity 21.5%; Pred. No. 6.2e-10;  
Matches 151; Conservative 310; Mismatches 238; Indels 4; Gaps 2;

Oy 3299 acataatattaaaatatagtaataataagtgtattaactttaattacagcataaat 3358  
| | : | :::: | | : || : || | : | : || | : | :|| | :  
Dh 398 ATATAATTTTTTTTAAWAAAAAATAATTAAWAAAAAATAATTAWAAAAAATAAT 457

Qy 3359 taaattttgaatcaattaatttttattctattattttaatttaatttagctatttttc 3418  
 : ||:|| |:|| || || |||: | || |:| |:| : | : ||::||  
Db 458 ATATATTTTAAWTAATAAAAAAAAAATTWTTTTTTWTATATATATATTTTAAW 517

[illegible][illegible]

**Qy** 3538 tctgattataattatgggtgggacacaatcgctttccactaaatattttaactatgatatta 3597  
:: :: | : : : | : : : : : : : : : : : : : :  
**Dh** 638 AAAAAAAAAAAGAAAAAGAAAAAGAAAGAAAATGTTTGGTAAGAAGATGTTTTTTTTTCACCAACTCTT 697

Qy 3598 taaattttttcaacatcgatatatttacttattaatacataatttatcataattttatgg 3657  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
nh 608 GGGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

[illegible]

Qy 3718 atgtacttttaggtaattttaagtactcttaaccaaacacaaaaattcaaatacaatgaa 3777  
 : ||| : | : ::||::| : : : : : : : : : : : : : : | : || :

Qy 3778 ctaaataagataatataacatcggaacattctacttgtaattttacattcccataattt 3837

          + + + + : + + + + : | :     | : + + : + + | : + : + + + + :

Qy 3838 tattatgaaantaatcttatattactgaactaaatgttgccacaattatta---tct 3894  
::: : : | : : : : : : : : : : : : | : : | : : ||:

Qy 3955 tgtatatttacgtaaaaatatttgacatagattgagcaccttc 3997  
: |: :::: ::::|:: : :: :| : :::

RESULT 4

|            |                                                                                                                                    |         |     |     |             |
|------------|------------------------------------------------------------------------------------------------------------------------------------|---------|-----|-----|-------------|
| LOCUS      | CNS0167M                                                                                                                           | 1201 bp | DNA | GSS | 26-JUL-1999 |
| DEFINITION | Drosophila melanogaster genome survey sequence T7 end of BAC BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit fly) |         |     |     |             |

ACCESSION AL106396  
VERSION AL106396.1 GI:5621701

KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Bukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefr@genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelBAC11.

```

FEATURES             Location/Qualifiers
    source            1..1201
                        /organism="Drosophila melanogaster"
                        /plasmid="pBeloBAC11"
                        /db_xref="taxon:7227"
                        /clone_lib="DrosBAC"
                        /clone="BACW15M24"
                        /note="end : 7'"
BASE COUNT            323 a    87 c    79 g    551 t    161 others

```

Query Match 2.2%; Score 121.6; DB 123; Length 1201;  
Best Local Similarity 37.4%; Pred. No. 3e-09;  
Matches 280; Conservative 101; Mismatches 368; Indels 0; Gaps 0;

**OY**    3219    aatgtaatatatttatttttaaataaaattgtttatttagattttaaatatttgagca    3278  
               |:| :| :|| | || | :| || :| :| | |  
**NB**    1193    wtatwttwtatataaaaataaaaaataataatgatataawnaaaanaantataaaaaaaa    1134

[illegible]

Qy 3339 ttaaattacaagcataatattaaatttgatcaattatattttatttatttttaa 3398  
:: ||: | |:: | : | || | | |||| | | : | : |

Qy 3399 ttaatttagtctatttttcaaaaataaatttaaatctaaataaaaaataatttttctcta 3458  
|| ||| | : | | : || || || | : : || ||

Qy 3459 atgttgaacaaactcatgttatacttcaaaattataagttattatattaccttgatgatt 3518  
| ||| : | : : | | : |||| : : : ||| : || |

Qy 3519 tatttattagtatattattctgattataattatggtgggatacaatcgctttccactaa 3578  
| : | : || | |||| : | || : : | :

Qy 3579 atattttaactatgattataaatttatttcaacatcgatatatttacttattaatacata 3638  
:||| ||| | :| ||| | | :| | | | | ||| :| :

Oy 3639 atttatcataattttatggaattgagaccaagaacattaagagaaacaattctataac 3698  
::| | ::|| ||:|| |||| || : ::|| ||

Qy 3699 aaagacaatttagaaaaaatgtacttttaggtaattttaagtactcttaaccaaacaca 3758

Qy 3759 aaaattcaaatcaaatgaactaaataagataatataacatacgaacatcttacttgtaa 3818

```

RESULT 1
CNS00EVL
LOCUS      CNS00EVL      1101 bp      DNA      GSS      04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit

```

RESULT 15  
 PCT-US92-00018-1/c  
 ; Sequence 1, Application PCT/US9200018  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hoffman, Stephen L.  
 ; APPLICANT: Charoewit, Tupin  
 ; APPLICANT: Hedstrom, Richard  
 ; APPLICANT: Khumthit, Sirin  
 ; APPLICANT: Rogers IV, William O.  
 ; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
 ; TITLE OF INVENTION: Immunogen and gene encoding  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: A. David Spavack  
 ; STREET: NMRDC Building 1 7-12 National Naval  
 ; STREET: Medical Center  
 ; CITY: Bethesda  
 ; STATE: MD  
 ; COUNTRY: USA  
 ; ZIP: 20814-5044  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.24  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/00018  
 ; FILING DATE: 19920101  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:

```

: NAME: Sperack, Avram D.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-6759
: TELEFAX: (301) 295-4033
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4673 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: N
: ANTI-SENSE: N
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium yoelii
: STRAIN: 17X(NL)
: DEVELOPMENTAL STAGE: erythrocytic stage
: TISSUE TYPE: Blood
: CELL TYPE: erythrocytic stage
: IMMEDIATE SOURCE:
: LIBRARY: Py-lambda/ttl1-2-7 kb genomic expression
: CLONE: Py10.1111
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 718..3195
: OTHER INFORMATION:

```

```

Query Match      1.44; Score: 79; DB 6; Length 4673;
Best Local Similarity 44.84; Pred. No. 64e-06;
Matches 354; Conservative 0; Mismatches 395; Indels 8; Gaps 3;

Qy 3220 atgatactatataattaaataaaattatgttatttagattcttaatttttggagcat 3279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3932 AAGGATGATATACTCTAAAAAATCATGAGCCAACTCTAGATATCTCCTTTTAAACAT 3873
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3280 tccatctactaatcttgcgaactaaattaaataatgataataaagtgaattaaactt 3339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3872 CCACTCATTTTTTTTATCACTACTTTTTAGTCATAAAACTAAGACCATATTTTATGT 3813
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3340 taastataacgaataattatasaattttgaactaaattttttattctattatttaaa 3399
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3812 TTGTGTAATTTTAAAAAATACACAAATTTGTTTATTCTTTTTTTCGATAATATT---GG 3756

Qy 3400 taattttgtactttttttcaaataaaatttaactcaataaadaatttttccctaa 3459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3755 TGGATTGTCTATTTTTTGAAGATCATTTTATATGTATATCACTTTTTTGTTTAT 3696
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3460 tgttgaaacactcctgtgtatacttcaaaattataagatttatatttacctgtgattt 3519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3695 TCAATATTATTTTGAAGATGTAATACCTGTAANAATACAACTCAAAAATGTTATA 3636
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3520 atttattagtataattatctgtatataattatgttggaatacaactgttccactaaa 3579
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3635 ACTTTTAAGTTTTTTTTTTTTTTTTTTTTTTTTATATAGAATATTGTAACCATTAAA 3576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3580 tatttttaactatgattataaattttttcaactcgtatatttactattataacataa 3639
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3575 TATTGGCATATAAATAATATATATATTAATAC--GAGACAACTAGTTAAAAAAAATAT 3518
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3640 ttatcataatttttggaaattgagacagagaacataaagagaacaattctataaca 3699
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3517 TTAaaaaaaAACTGTTAAAAAATAATGAAATAATGAATATAAATTGATGATAAAT 3458
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3700 aagacaattttgaaaaaaatgtacttttagtataattttagactcttaacaaacacaa 3759
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3457 CATAAAA---AGAGACAAATTTATCAAACTGTTAATTTTAAATAACTTTAATCATAAAA 3401
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3760 aaattcfaatcaattgaactaataatgataataataacatacggacaactctacttgta 3819
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3400 AAAAAGAGACAAAAACAGAAAAAAGATAAAGTGTAAAGAAACAAACAACTTTAAAGA 3341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Qy 3027 gttgctgtctc-----taccacaagatgatacaagtgtttagctgcttccaaagaca 3079  
Db 1012 ATGTAAATAAAGCTGTTTATACCTCTTTTATAAAAAAATAATATATATATAATAG 953

Qy 3080 atttttgcccaataatgataaagcaattgttttagttcnaactgcacagaataatgtt 3139  
Db 952 GGAATGCTTACAAGATATAAAAGTAAATATACAGATATATATAGAATATCTATAGTTT 893

Qy 3140 aaaaagaattaaaataagttgctgctgcacacacacaa-----aaaaaacta 3189  
Db 892 ATATAGGAAACCAATAGATATATATATCTTTTATATACACACTTTGCGATTTAAAGATA 833

Qy 3190 atgttggttggtgaattttattattacggaatgaattattatttttaaaataaattat 3249  
Db 832 AACTCTTTAAGACCTATGATTCAGAGAAATCCCAATAATATATATATATATATAT 773

Qy 3250 gttattagattcttaataatttggaagcttccatctataatttctgaacataatata 3309  
Db 772 TATATATATATATATATATTTTTCCTATTTCTTTTTTTTATATACATATATAT 713

Qy 3310 aaatatagataataaagtttaatttaacttaatttaaacagcagataataaattttgaa 3369  
Db 712 TATGTTAAAAATTTTATAAATTTACATATACAGTGTATTTTCATATGAAATTTTT 653

Qy 3370 tcaatttaattttattctattatttttaatttaattagtcatttttcaaaataaaatt 3429  
Db 652 TTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTAGTAGAATACTATTTAT 593

Qy 3430 taacttaataaaaataattttcttaattggtgaacaacactgattatacttcaaaa 3489  
Db 592 AACTTAGCAAAAAAATAATTAATGAATAAATTAATAAATAATATAAAATAAATTA 533

Qy 3490 ttatagattattattacctgtgatgtttattattagtagtataaattctgtatataat 3549  
Db 532 TAGGACACCAATATTTGGGGAGATTATATATATGTGTAAATATATAGGATGGTTTT 473

Qy 3550 tatgtgggatacaactgcttccactaaatttttaactatgattataaattttatttc 3609  
Db 472 ATAGAGGTGTAACATATATTAAATGTGTACACATCAAAATATTAATATGTGATCTATA 413

Qy 3610 aacactgtatttttacttattataacataaattttatcataattttt--ggaattgagac 3667  
Db 412 TTAATTGGAATATATATAATAGTTTATATTATTCAGTAATTTTATATAAAGATGACAC 353

Qy 3668 caagaacaactaagaagcaaatctcatacaacaagacattgaagaaaaattgtctttt 3727  
Db 352 ACAACATATATATATATATATATATATATATATATATATATATATATATATATAT 293

Qy 3728 aggttaatttaagtcctcttaaccacacaaanaattcaaatcaaatgactaaataaga 3787  
Db 292 ATGTATGTGTTATTAATATAAAGAAAAAATAAATATATATATATATATATATATAT 233

Qy 3788 taataataacatacgagcaacttacttgytaacttaccatccocataaattttataga 3847  
Db 232 TATATAGAAATATCAATATATATATCTACATCAATCAATATAAATATTTTATCTTTTAT 173

Qy 3848 aataactattataact 3864  
Db 172 TTTTATTATTTATTT 156

RESULT 12

US-08-883-795A-36/c

; Sequence 36, Application US/08883795A

: Patent No. 5985607

: GENERAL INFORMATION:

: APPLICANT: Delcuve, Genevieve

APPLICANT: Awang, Gregor

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator

```

; NUMBER OF SEQUENCES: 39

```

CORRESPONDENCE ADDRESS:

```

ADDRESS: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3T2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Michelle
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062

TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 665 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: rh 32

```

Query Match 1.5%; Score 82.4; DB 4; Length 665;  
Best Local Similarity 51.0%; Pred. No. 1.2e-06;  
Matches 252; Conservative 0; Mismatches 231; Indels 11; Gaps 2;

[illegible]

APPLICANT: Chang, Andy C M  
APPLICANT: Williams, Keith L  
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular  
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris  
STREET: One Liberty Place 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/867,106  
FILING DATE: 19920625  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PJ 7187  
APPLICATION NUMBER: PCT/A090/00530  
FILING DATE: 02-NOV-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Feeney, Joanne Longo  
REGISTRATION NUMBER: 35,134  
REFERENCE/DOCKET NUMBER: RICE-0002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5852 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
US-07-867-106-2

Query Match 1.54; Score 85; DB 1; Length 5852;  
Best Local Similarity 46.14; Pred. No. 6e-07;  
Matches 403; Conservative 0; Mismatches 460; Indels 12; Gaps 3;

Qy 3120 actgctcacagaaataagtttaaaataaataaagtggtggtgcacacacaa 3179  
Db 1381 AGTAAGACATAAATATGATAAACCACAAATGATTAAGATAAGTATATGTTGT 1440  
Qy 3180 aaaaaaactagttggttggtgaattttatattacagtgtaattatatttttaa 3239  
Db 1441 ATTATATATATAGTTAGTTTAAAGTTTATATCATTTTTTAAAAAGTAAATGTTG 1500  
Qy 3240 ataaatattgtttatttagattcttaattttggagatccatactataatttggtaa 3299  
Db 1501 AAAAAAATTTTTTTTTTTTTTCAACGGGAGAGATTAATATCATATGATCAAAAT 1560  
Qy 3300 cataataataaattagataataaagtgtaatttaattacagcataatt 3359  
Db 1561 TAAAGGTATTAACAATATGTAATAAATTAATAAACTAACCTGATATATATCTTC 1620  
Qy 3360 aaatttgtagaataatttttttttttttttttttttttttttttttttttttt 3416  
Db 1621 CCCCTCTTTTTTTTTTTTTTTTGTCTAGACACTTTTTTTTTTGTCTAGACACTTTT 1680  
Qy 3417 tcaaaataaatttaattctaaataaaatttttttttttttttttttttttttttttt 3476

Db 1681 TTTTAAAAAATAAATAAATGTAATAACTATTGTGACATCAITTTTCTAGTT 1740  
Qy 3477 ttatacttcaaaattatagattatattaccttgatgattatttagt--atat 3534  
Db 1741 TTTTTTATGATATATAAATAAATGCTTATGATATATATCTAATTTATATGATT 1800  
Qy 3535 aattctgattataattatgtgggatacaactgctttccataaatttttaactatgat 3594  
Db 1801 GAATATATTTTAAATTTTATAAATCTACTTTTTTTTTTTTCTTTTTTTTAA 1860  
Qy 3595 ttataaattttttacacatgataatttttactataataaattttatcataattta 3654  
Db 1861 TTTTAAATTTTTTTTTTTTATAGATCTCATTAATAAATCAATTAATAATAAAGT 1920  
Qy 3655 tgaatttgagaccagaacattaaagacaataattctatacaagacacattagaaa 3714  
Db 1921 TATTTTTTAAATATGCAAAACTATAAAACTAATATGATTTAACCACTTTTTCTATT 1980  
Qy 3715 aaattgacttttagtgtaatttttagtactcttaaccaacacaaatttcaaat 3774  
Db 1981 TCTTTTTTTTTTTTTTTTTTTTTTACTTGAATAAATAAATAAATAAATAAATA 2040  
Qy 3775 gaactaataagataatacaactacagcaacttactgtatatt-----acatt 3827  
Db 2041 AAACCTCATTAATAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100  
Qy 3828 ccaataattttatattgaaataattctataattactgaactaaattgttgcacaa 3887  
Db 2101 AAAATTTTATCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
Qy 3888 attatctaaataagaaacacacttaattttataacatttttccataatttgaagat 3947  
Db 2161 TTATCAGTTTGGCATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2220  
Qy 3948 tatattttgtattattacgtaaaaattattgacat 3982  
Db 2221 TTTTTTTTTTAAATTTCTTTTTTTTTTTTTTTTAT 2255

RESULT 10  
US-08-883-795A-36  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcove, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3T2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelser, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: WIH21.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8350  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 1.84; Score 100.4; DB 4; Length 19124;  
Best Local Similarity 44.19; Pred. No. 1.4e-09;  
Matches 877; Conservative 0; Mismatches 1091; Indels 22; Gaps 10;

Qy 1970 atigaacgctttaagaatttttactactgcaaatcgaataagtgaaattgttttttag 2029  
Db 4660 ATTCATATTTAGAGATTATGTAATGTTTATGTTATGATGTAATATATATATTAATT 4719  
Qy 2030 aaagattaaataagttagtattacagcttttttagtttggttggtggaagaaatgatgt 2089  
Db 4720 GTTTTTAGTATGATGATGATCTAATAATATATTCATATGATGTCATAGTGCAATGA 4779  
Qy 2090 ttttgacaataatttttgacaataaagtttttagtttcagggaataacggaaatattctc 2149  
Db 4780 ATATATAATATGATATTTATATATATGTTATATATAATTAAGTAACACACAG-ACATTA 4838  
Qy 2150 tcttttttgtaaaattactaattgcaagaacaaacaagctttggggagcaaaatactcta 2209  
Db 4839 TATATAGTAATAAATAGAGAAATATATATATTTTATGTTATATATATAGTTATATA 4898  
Qy 2210 gottttaagtagtcggtgaactctcaaaactcggctcaaaacttcctaggtgagttgtgtg 2269  
Db 4899 AAGGGAAATTCATATATTTTGAATTTTGTATATGATATAGTTATAGTTAAAA 4958  
Qy 2270 tgcctacgtatgaagtctatagaacttcctgacaaacagcatgagctcagggtcgaa 2329  
Db 4959 AAAAAAAGAACAGAACAAATGGAAGCATAAAAATGTTCTGTATAGGATAAAA 5018  
Qy 2330 tctacaacttttcccttttcttcaataacatagtgttgctgaagtttcgactataat 2389  
Db 5019 TATATATATAAATGTTTATTTATCTTAAAAAGGTTCCCTATATATACATAAAAAAA 5078  
Qy 2390 aatttattacagtttatcaatttcaattcattatcatatcatttataaataagctca 2449  
Db 5079 TTGTGCCATTTTATAAATTAATACATACATTACATATGAATTCGATTTGTGTGTTT 5138  
Qy 2450 gtccaattcagtttttogaagttccccaanaatttgaaatttataaatttattccctaa 2509  
Db 5139 TTTTGATGAATATATGAGCTAATATATATATATGTAAGTGCTGTATATATAATAATA 5198  
Qy 2510 aacggaaatgattatcttttcaaatgaagtttcaatttcaattcagtttcaattc 2568  
Db 5199 ATTTTATTTAAAAAATGAAAAATGAAGATAAATATCTCGATTTTGTATGTTCCATAGC 5258  
Qy 2569 atctttttataactctctattatctataaattcaataaatttcaaatatttttgaaatat 2628  
Db 5259 TTAATATAATATGGACTCATATATATATATATATCTTTACACAGTAATAGTAA 5318

Qy 2629 ttacacttttagtccctaaagttccaacaataaatttccactttagaataatcatTTTT 2688  
Db 5319 ATATATTTTAACTCTTAATAGGAAATAAATAATATAAATAGAA-...TACTGAATA 5375  
Qy 2689 cacctctaaagctcaaaatttcaacaataagcaaaatttccactgattagttgaagcaagt 2748  
Db 5376 TAGTCATATATATACATTTTTTAAAAATTAACATATATACAAATAGCAATCATGTATTA 5435  
Qy 2749 ttgttgcttccaacaataaaaaatttcaaaaaaaacaacttaaaactctttatcaat 2808  
Db 5436 TAGAAATATAGAAATTTAATAATTAAGGATAAATATATATTTAAAAATATATTTTTT 5495  
Qy 2809 ttgaacacaagaagcttgccgaatgctgaagcttcaaaatgctctttttgtttctttt 2868  
Db 5496 ATGTCAATTTATGTTTATATATATATATTAACATGATTA-GTTTTTGAAAAATATTTA 5554  
Qy 2869 ttgtgcaaacgctggaggaagaggggaagaagattgacatatTTTTtattatgttt 2928  
Db 5555 AATATCATATATAATAATAAATTAGTTAAATAATATAGTTTTCATACAAATACTACT 5614  
Qy 2929 taacataataataataaatttaataataatttactgttggaatgagcaagtgggag 2988  
Db 5615 TATAGATATATCATATAATATATATATATATATTTATGTTTTGTTTGGGTGAT 5674  
Qy 2989 atacgtaaagtatttttaacattatattcttttccaagcagttgctggttaccocaaagt 3048  
Db 5675 ATAGGCTTATAGTATATATGTTGGTTCATATATATTTATATGGAATAGATCATAT 5734  
Qy 3049 gatcaaa-----gtttgagctgccttcaatggccaatttttgcccatgaatgataaa 3101  
Db 5735 AAGTAATATATTTTATTGTTATATGTCGTGTTAAGATAGATGCAATACAGTTAAG 5794  
Qy 3102 ggcactttgttttgactgactgctcaagaataatgttcaaatgaattaaataaggtg 3161  
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTTTTGTACATATATATAAAAAATAGATACTACATA 5854  
Qy 3162 gctgtgctacacacacacacacacacacacacacacacacacacacacacacacacac 3221  
Db 5855 TGCATATATACAGAAATAATTTGTATAAAATATATATATATATATATATAAAGACAT 5914  
Qy 3222 gtaattatatttttaaaataaattatgtttttagattcttaatttttggagcattc 3281  
Db 5915 -TAAACATATCTAATAGGTAAATAGTTTTTATATATATATCTCTTTTATATATATTTT 5973  
Qy 3282 cactactataatttgcataataattcaaatatagtaataaagtgttaatacttta 3341  
Db 5974 TTTTGTTTTACTCTGTGCTGTTTTTTTGTATATATATATACAAATATATAAACAATA 6033  
Qy 3342 aattcaagcataatataaattttgaatcaataaattttattttattattatttaetta 3401  
Db 6034 TCAGTATTTGGATATAAATAAATTTTCTTCAATATGATCATATATATATATATATA 6093  
Qy 3402 atttagctcttttttccaataaataatttaactcaataaataaattttt---octta 3458  
Db 6094 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6153  
Qy 3459 atgttgaaacacactgcttactactcacaattataagattattatttacccttgatgt 3518  
Db 6154 ATTTTATATATTTTATGATATACCTTTAAGATATATTAATATATATATATATATATAT 6213  
Qy 3519 tattttattagatttaattctgattataattatgtgggatacaactgcttccactaa 3578  
Db 6214 ATGTATTTATATATATACAAATTTTTCATTTATATATATATATAGAACTAGAACATTT 6273  
Qy 3579 atatttttaactagattt-tataaattatttccaactgattatttactatttaatacat 3637  
Db 6274 ATTAATACCTATATTTGAATATATATATATATATATATATATATATATATATATATAT 6333  
Qy 3638 aattttatcaatttttttggaattggagccagaacacattagagacacacattctata 3697  
Db 6334 ATTAATCAATAAATTTTGAATTCATAAATGATGAATATACATAAAAAATACACAA 6393

RESULT 5  
 US-08-787-335-18  
 ; Sequence 18, Application US/08/87335  
 ; Patent No. 5961834  
 ; GENERAL INFORMATION:  
 APPLICANT: John, Mallyakal E.  
 APPLICANT: Umbeck, Paul F.  
 APPLICANT: Brill, Winston J.  
 TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS  
 TITLE OF INVENTION: POR ALTERED FIBER  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Quarles and Brady  
 STREET: P.O BOX 2113  
 STREET: FIRST WISCONSIN PLAZA  
 CITY: MADISON  
 STATE: WISCONSIN  
 COUNTRY: U.S.A.  
 ZIP: 53701  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Macintosh  
 SOFTWARE: Microsoft Word 4.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/787,335  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/530,797  
 FILING DATE:  
 APPLICATION NUMBER: US 07/253,243  
 FILING DATE: 04-0CT-88  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Nicholas J. Seay  
 REGISTRATION NUMBER: 27,386  
 REFERENCE/DOCKET NUMBER: 1122990245  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1283 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: no  
 ANTI-SENSE: no  
 ORIGINAL SOURCE:  
 ORGANISM: Gossypium hirsutum  
 STRAIN: Coker 312  
 DEVELOPMENTAL STAGE: 15 day old fiber cells  
 TISSUE TYPE: fiber cells  
 IMMEDIATE SOURCE:

```

; LIBRARY: CKFB15
; CLONE: E9
US-08-787-335-18

```

Query Match 4.8%; Score 265.4; DB 4; Length 1283;  
Best Local Similarity 84.3%; Pred. No. 9.4e-39;  
Matches 316: Conservative 0; Mismatches 46; Indels 13; Gaps 1;

QY 4132 aataacagctttttttttctattgataaaccaatgctctatagacttgcaccccttc 4191  
Db 13 ACTAAAAATCTTGGCTTCTATTCTTGTGAACCAAGCTCTAAACTTTGTCACTCTTC 72

QY 4192 tctcttttccacttttactctaaagtgctctactagtcgacggtgacacacttttgc 4251  
Db 73 TCTCTTTTCACMTTTCATCTACTGCTCACTATATTGGTAGTCAACCGCTCTGC 132

QY 4252 gcagcgctctgagttttatctgagacacagcaacctctcagagctcccaaatgtgct 4311  
Db 133 TCAGCGGCTGCMATTATTTCGACGACCAACCAACCTCTCAGAGCTGCACATGTGCT 192

QY 4312 tcaaaatacgaagcagcagagctctgaatacgaagaacgaatacaaaacgccaagta 4371  
Db 193 TCATAATACGAAA-----AGCAACAAGAGTCTGAATCAAAACAACCAATA 239

QY 4372 tcacgaagagtactcaaaacttgagagcgtcgaaatgcacaaaggagaanaaacaac 4431  
Db 240 TCAGCAAAAGTACCAAAATACGAGAACTTGAATGCTACAGAGGAGGAAAAACAACAA 299

QY 4432 ctgcacacagcatgaagagtaacacagctacacgaatcaaaaggacaaaaagataga 4491  
Db 300 CTCGACACATCATGAGAGATACACAGAGTCACGCGAATCGAAGGAGCACGAAGATACGA 359

QY 4492 gaagagaatacttoga 4506  
Db 360 TAAAGAAAACCGGA 374

```

1  RESULT 6
2  US-08/487-826B-13/c
3
4  ; Sequence 13, Application US/08/487826B
5  ; Patent No. 593827
6
7  GENERAL INFORMATION:
8
9  ; APPLICANT: Sin, Kim L.
10 ; APPLICANT: Chitnis, chetan
11 ; APPLICANT: Miller, Louis H.
12 ; APPLICANT: Peterson, David S.
13 ; APPLICANT: Su, Kin-zhaun
14 ; APPLICANT: Wellens, Thomas E.
15
16 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
17 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
18
19 ; NUMBER OF SEQUENCES: 45
20
21 ; CORRESPONDENCE ADDRESS:
22
23 ; ADDRESSEE: Knobbe Martens Olson & Bear
24 ; STREET: 620 Newport Center Drive 16th Floor
25 ; CITY: Newport Beach
26 ; STATE: California
27 ; COUNTRY: US
28 ; ZIP: 92660
29
30 COMPUTER READABLE FORM:
31
32 ; MEDIUM TYPE: Floppy disk
33 ; COMPUTER: IBM PC compatible
34 ; OPERATING SYSTEM: PC-DOS/MS-DOS
35 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
36
37 CURRENT APPLICATION DATA:
38
39 ; APPLICATION NUMBER: US/08/487,826B
40 ; FILING DATE: 10-SEP-1993
41 ; CLASSIFICATION: 435
42
43 ATTORNEY/AGENT INFORMATION:
44
45 ; NAME: Israelsen, Ned
46 ; REGISTRATION NUMBER: 29,555
47 ; REFERENCE/DOCKET NUMBER: NH121.001CP1
48
49 TELECOMMUNICATION INFORMATION:

```



```

: TISSUE TYPE: fiber cells
: IMMEDIATE SOURCE:
: LIBRARY: CKFB15
: CLONE: E9
US-07-685-970A-17

Query Match          4.88; Score 265.4; DB 1; Length 1283;
Best Local Similarity 84.3%; Pred. No. 9.4e-39;
Matches 315; Conservative 0; Mismatches 46; Indels 13; Gaps 1;

```

Qy 4132 aataacagttcttttctttctatttgattaaccatggctcatagcattcgccacccttc 4191  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 13 ACTAAAAATCTTTGCTTTCATTTTGTAAACCATGGCTCATAACTTTTGTCATCCTTC 72

Qy 4192 ttctttttccaaacttttactcataagtgtctcactagtgaccggtagccacactgttgg 4251  
Db 73 TTCTTTTCCAACTTTTACTCATTACTCTCTCACTAATAAATCGGTAGTCACACCGCTCTG 132

Qy 4252 gcagcggctcgacgtttattcgcagacacaaagcaacctcatcagagctccacaaattggct 4311  
pb 133 TCAGGGGCTCGACGTTTATTTCGCAGACACAAAGCAACCTCATCAGAGCTCCACAAATTGGCT 192

Qy 4312 tcaaaatacgaagcagcagagctcgaatacgaagccagaatacaaacagccaaagta 4371  
|||||  
pb 193 TCAAAATACGAA-----ACCAACAGAGCTCTCAATACAAACAAACAAATA 239

Qy 4372 tcacgaagagtactcaaaactgagaagcctgaatgcaaaaggaggaaaaacaataaac 4431  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
pb 240 TCACGAAAGAGTACTCAAAACTGAGAAGCCTGAATGCAAAAGGAGGAAAAACAATAAAC 289

Qy 4432 ctgcaaacagcatgaagagtaccacgagtcacacgaatcaaaggagcaaaaagagtacga 4491  
pb 300 CTGCGAAGATGTCACGCTGGTCCGTGTTGCGGCAATGCGACGACGCGCATGCGCGTA 350

Qy 4492 gaaagaaaatctcga 4506  
||||||| | |||  
ph 360 gaaagaaaatctcga 374

```

1  RESULT 2
2  US-08-298-687A-17
3  ; Sequence 17, Application US/08298687A
4  ; Patent No. 5521078
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: John, Maliyakal E.
9  ; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
10 ;
11 ; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
12 ;
13 ; NUMBER OF SEQUENCES: 33
14 ;
15 ; CORRESPONDENCE ADDRESS:
16 ;
17 ; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
18 ; STREET: P.O. Box 2113, First Wisconsin P.A.
19 ; CITY: Madison
20 ;
21 ; STATE: Wisconsin
22 ;
23 ; COUNTRY: USA
24 ;
25 ; ZIP: 53701
26 ;
27 ; COMPUTER READABLE FORM:
28 ;
29 ; MEDIUM TYPE: Floppy disk
30 ;
31 ; COMPUTER: IBM PC compatible
32 ;
33 ; OPERATING SYSTEM: PC-DOS/MS-DOS
34 ;
35 ; SOFTWARE: Microsoft Word
36 ;
37 ; CURRENT APPLICATION DATA:
38 ;
39 ; APPLICATION NUMBER: US/08/298,687A
40 ;
41 ; FILING DATE:
42 ;
43 ; CLASSIFICATION: 800
44 ;
45 ; PRIOR APPLICATION DATA:
46 ;
47 ; APPLICATION NUMBER: US 07/617,239
48 ;
49 ; FILING DATE: 21-NOV-1990
50 ;
51 ; PRIOR APPLICATION DATA:
52 ;
53 ; APPLICATION NUMBER: US 07/253,243
54 ;
55 ; FILING DATE: 04-OCT-1988
56 ;
57 ; ATTORNEY/AGENT INFORMATION:
58 ;
59 ; NAME: Seay, Nicholas J.

```

```

REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum
STRAIN: Coker 312
DEVELOPMENTAL STAGE: 15 day old fiber cells
TISSUE TYPE: fiber cells
IMMEDIATE SOURCE:
LIBRARY: CKEF15
CLONE: E9
US-08-298-687A-17

```

Query Match 4.8%; Score 265.4; DB 1; Length 1283;  
Best Local Similarity 84.3%; Pred. No. 9.4e-39;  
Matches 316; Conservative 0; Mismatches 46; Indels 13; Gaps 1;

Oy 4132 aatacagcttcttttctttctatttgattaacatgggtcatagcattcgccaccttttc 4191  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 13 ACTAAAATCTTTGCTTCTATTTTGTAAACCATGGCTCATAACTTTTGTCATCCCTTC 72

Qy 4192 ttccttttccaacttttactcataagtgtctcactagtacggtagccacactgtttcg 4251  
Db 73 Ttcccttttccaacttttactcattactgtctcactaataatcggtagtcacaccgctctcg 132

Qy 4252 gcagcggctcgacgtttattcgagacacaagcaacctcatcagagctcccaaatggct 4311  
Db 133 TCAGCGGCTCGACATTTATTTCGAGACACAACCACTTCATCAGAGCTGCCCAATGGCT 192

Oy 4312 tcaaaatacgaagcacgagagtctgaatcgaagaacccagaatacaaacagccaaaagta 4371  
          |||||         ||| || |  
pb 193 TCAAAATACGGA-----AGCCAAAGAGTCTGAATCAAAACACAAAAA 238

Oy 4372 tcacgaagagtactcaaaacttgagaagcctgaaatgcaaaaggagggaaaaacaataacc 4431  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 4432 ctgcaaacagcatgaaggataccacgagtccacgaatacaaaggagcaaaaagagtaca 4491  
|||||  
rh 300 GTGGAAGATGGTGTCTGCACGTCAGCGCAGCCCAATGCCAACCACGCGCATCGTGACA 358

Qy 4492 gaaagaaaatctcga 4506  
 ||||| | |||  
 sh 368 gaaagaaaatctcga 374

RESULT 3  
US-08-530-797-18  
; Sequence 18, Application US/08530797  
; Patent No. 5597718

GENERAL INFORMATION:  
APPLICANT: John, Maliyakal E.  
APPLICANT: Umbeck, Paul F.  
APPLICANT: Brill, Winston J.  
TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS  
TITLE OF INVENTION: FOR ALIRED FIBER  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles and Brady  
STREET: P.O. BOX 2113  
STREET: FIRST WISCONSIN PLAZA  
CITY: MADISON

Db 5555 AATATCATATATATATAATTAATTAAGTAAAAATATAGTATTTTCATCAAAATACTAACT 5614  
Qy 2929 taacataataatttaatttaatacttaactcttggtgaagtgacagtgaggag 2988  
Db 5615 TATAAGTATATCATATATATATATATATATATATATATATATATATATATATATAT 5674  
Qy 2989 atacgaagatttttaacattatctttttgcaagcagtggtgctaccccaagtg 3048  
Db 5675 ATAGGCTATAGTATATATGCGTTGTCATTATATATATATATGTAATAGATACATAT 5734  
Qy 3048 gatcaaa-----gtttgaagctgcttcaatgagccaaatttttgccataatggataaa 3101  
Db 5735 AAGTTATATATTTATTTGCTATATGCTGTTGTAAGATAGATATGCATTACAGTTAG 5794  
Qy 3102 ggcatttggttagttcaactgctcagcaataatgtttaaagtaaatgaagtg 3161  
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTTGTACATATATATAAAAAATAGATACTAACATA 5854  
Qy 3162 gcttgctcagcagcaaaaaaactaagtggttggtgaattttatattacggat 3221  
Db 5855 TGCATATACAGAAATATATTTGTATAAAATATATATATATATATATAAAGCAT 5914  
Qy 3222 gtaattatatttttaataaaattatgttattagattcttaattttgagcattc 3281  
Db 5915 -TAAACTATACTAATAGGAATTAGTTTTATATATCATCTTTTATATATATAATTT 5973  
Qy 3282 catactataatttctgaacataatattaaatatagtaataataagtgtaattacotta 3341  
Db 5974 TTTGTTTACTTCTCTGCTGCTTTTTTGTATTATATATAACAAATATAAACATA 6033  
Qy 3342 aattcaagcataataattttgaatcaattattttattctatttttaatta 3401  
Db 6034 TCAGTATTGGAAATATAAATAAATTTATTCACATATGCAATATATATATATATA 6093  
Qy 3402 atttagtctatttttcaaaaaaaatttaaatctaaataaaataattttt---octta 3458  
Db 6094 TATATATATATATATATATATATATATATGATGATTATATCTATTTTATACATGC 6153  
Qy 3459 atgttgaacacactcatgttatacttcaaaattataagttattatttaacttgatgatt 3518  
Db 6154 ATTTTATATATTTTAGTATATACTTTAAAGATATATTAAATTTATATATAGTACATAT 6213  
Qy 3519 tatttatagattatttaattctgattataattatggtggatcacatogcttccacaa 3578  
Db 6214 ATGTATTATATATATACAAATATTTTCATTATATAATATATAGAACATGACATTTT 6273  
Qy 3579 atattttaacttgatt-tataaattttttcaacatoytatatttacttattaacat 3637  
Db 6274 ATTAATACATCATATTGAAATATATATTTAATATGATGATTTTACTATTTTTTAT 6333  
Qy 3638 aatttatcatattttatggaattgagaccagaaacattagagacaaattctataa 3697  
Db 6334 ATTATACATATAAATTTTGAATTCATAAATGCATGAATACATAAAAAATACACAA 6393  
Qy 3698 caaagcaaattagaaaaaat-gtacttttagttaatttaagtaactottaaccaaca 3756  
Db 6394 AACAAATGATAAACAATTTTTTATTAATATAATATATATATATATATATATATTTTC 6453  
Qy 3757 caaaattcaaatcaaatgaactaaataagataaataacatacaggaactcttactgt 3816  
Db 6454 CTGTTATTATTTATCATTTTTTTTTGATGCTATATATATATATATAATAAATATATA 6513  
Qy 3817 aattctacattcccaataattttatgaaataaactctattattactctgaactaagt 3876  
Db 6514 ATATATA---ACACAAAAATTAATATAATATAATATACTACTTTTAAATATAACACAA 6570  
Qy 3877 tgtccaaattattctcaataagaacaaacactaaattttataacattttttctat 3936  
Db 6571 TACAAGAAATGATATCTATATCAATATATATATATGAATATATAATATGATAGATAA 6630  
Qy 3937 atttgaaga 3946  
| | | | |

Db 6631 TATAGATAGA 6640

Search completed: September 3, 2000, 03:05:49  
Job time: 26295 sec

Qy 3026 agttggctggtctaccagaagatgatcaaaagtttgagctgcttccaatgagccaaatttt 3085  
 Db 6397 TGTGTGTGTATTTTTTATGTATTTATGTCATTTATGAATTCAAAATTTTATGTA 6138  
 Qy 3086 gccataatgataaaggcaatttggtagttcaactgtccacagataatgtttaaagt 3145  
 Db 6337 TAATAATAAAAAAAGTAAAAATACACATTAATAATATATATTCAATATAGGTAT 6278  
 Qy 3146 aaattaaataaagtgccctgtccacaccccaaaaaaaactaatgttgggtgtgaa 3205  
 Db 6277 TAATAAAATGTCATGCTTATATATTATATAAGAAAAATTTGTTATATATAAAT 6218  
 Qy 3206 tttttattacgaagtataattatattttaaataaaattgttttttagctctta 3265  
 Db 6217 ACATATATGCTACTATATAAATATAATAATATCTTTAAGAT-ATACATAAATATA 6159  
 Qy 3266 atattttgagcttccatctataattttoytaacataatataaaatagataata 3325  
 Db 6158 AAAATGTCATGATAAAAGTATAAAATCATACATATATATATATATATATATAT 6099  
 Qy 3326 agtgtaattacatttaaatcaacagataatttaaatttgaatcaattttttatt 3385  
 Db 6098 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 6039  
 Qy 3386 tctattattttaaatttttagctcttttttcaaataaaatttaacttaataaaaa 3445  
 Db 6038 ACTGATATGCTTTATATATTGTTATATATATAACAAAAAGACACAGAAGTAA 5979  
 Qy 3446 taatttttccatttggaaacactcgttatacttccaaattataagttattatt 3505  
 Db 5978 CAAAAAATATATAATAAAGAGTATATAATAAACTAATACCTATTAGTATAGT 5919  
 Qy 3506 taacttgtagattttatttagtatataattctgatatataa----toggggata 3561  
 Db 5918 TTAAATGCTTTATATATATATATATATATATATTATTAACAATATTATCTTGTA 5859  
 Qy 3562 caatcgctttccataaatttttaactatgattataaattttttccactgtatat 3621  
 Db 5858 TGCATATGTAGTATCTATCTTTTTTATATATATGACAAAAAATATATATAT 5799  
 Qy 3622 ttactttaatacataattttatcaatttttggaaattgagccagaacatttaag 3681  
 Db 5798 AACCCCTTAAGCTATGATATCTATCTTAACACAGCATATACACAAATAATATAT 5739  
 Qy 3682 agacaataattctatacaagacaaatttagaaaaaattgacttttaggttaatttaag 3741  
 Db 5738 ACTTATATGATCTATACATATAAATATATATATGAACACCCATATATATCTTA 5679  
 Qy 3742 actcttaacaaacacaaataattcaaatcaatgacttaataagataataaactag 3801  
 Db 5678 TTATATACCAACCAATCAAAAACACATAAATATATATATATATATATATATATAC 5618  
 Qy 3802 gaacatcttaattgtatcttacttccatattttatgatgaataaactctattatt 3861  
 Db 5619 TTATAGTATAGTATTGTTGAAATACATATTATTTAACTAATTTATATATATAT 5560  
 Qy 3862 actcgaaactaaattgttccaaattttatctcaataaagaagaaaaacttaatttat 3921  
 Db 5559 ATATTAAATATTTTCAAAAACTAATCATGTTATATATATATATATATATATAT 5501  
 Qy 3922 aacattttttctattggaagattattttgtatttattgaataaaatt 3976  
 Db 5500 GACATAAAAAATATATTTAAATATTTATATTATCTTAAATATATATATCTT 5446

RESUME 14  
 V22740  
 ID V22740 standard; DNA; 3701 BP.  
 AC V22740:  
 DT 28-SEP-1998 (first entry)  
 DE Babesia microti HWI-10 antigen sequence.  
 KW antigen; detection; diagnosis; vaccine; tick-borne disease;

HW differentiation; Lyme disease; ehrlichiosis; ss.  
 OS Babesia microti.  
 FH Key Location/Qualifiers  
 FT CDS 1210..2559  
 FT /tag= a  
 FT /product= antigen  
 FN EP-834567-A2.  
 FM 08-APR-1998.  
 PP 01-OCT-1997; 117067.  
 PR 24-APR-1997; US-845258.  
 PR 01-OCT-1996; US-723142.  
 PA (CORI-) CORIAX CORP.  
 PI Houghton R, Lodes MJ, Reed SG, Sleath PR;  
 DR WPI: 98-195465/18.  
 DR P-PSDB; W56290.  
 PT Polypeptides comprising Babesia microti antigens and their immunogenic fragments or epitopes - and related nucleic acid, vectors, transformed cells and antibodies, useful for diagnosis of PT infection and in protective vaccines  
 PS Claim 8; Page 32-35; 113pp; English.  
 CC The sequence is that encoding a polypeptide comprising at least one antigenic portion of a Babesia microti antigen. It can be used to diagnose B. microti infection by detecting specific antibodies in usual immunoassays. Infection can also be diagnosed using:  
 CC (a) primers or probes derived from the coding sequence, in standard amplification or hybridisation tests, or (b) using antibodies to detect the corresponding antigen. It is also useful in vaccines to protect against infection, especially when formulated with an adjuvant. The new diagnostic methods allow rapid differentiation between B. microti infection and other tick-borne diseases (Lyme disease and ehrlichiosis) that CC have similar symptoms but require different treatments.  
 SQ Sequence 3701 BP; 1458 A; 457 C; 492 G; 1294 T;  
 Query Match 1.88; Score 101; DB 1; Length 3701;  
 Best Local Similarity 47.4%; Pred. No. 6.3e-06;  
 Matches 398; Conservative 0; Mismatches 43; Indels 6; Gaps 3;  
 Qy 3208 ttatatccagtggaatttaatttatatttttaaaataaattatgtatttagattctta 3267  
 Db 369 TTATATTCATGTTGGTATAATAATAAGTATATATAGTTTGTGATTTGATGATATA 428  
 Qy 3268 attttgagacttccatcataaatt----toytaacataatataaaatagataata 3323  
 Db 429 AATTAGAACAGATATATTAATAATTCAATATTATATTAAATTTTATATATATGATT 488  
 Qy 3324 aaagtgaattta-actttaaatcaacagataattataaatttgaatcaatttaatttt 3382  
 Db 489 TATTGATATTATATATATATATATATTGTTATGATCATTTAATGATATATATCAAT 548  
 Qy 3383 attttctatttttaatttaattgtctattttttcaaataaaatttaaatctaaata 3442  
 Db 549 CCATATATATATATAAATGATATATAAATTAATTTATGGCATATTACATTTATAAT 608  
 Qy 3443 aaataatttttccatttggaaacactcgttatacttccaaattataagattatt 3502  
 Db 609 AATATATATTATGTCATATGACATATATATATATATCCCATCATGATTGTAAGTAA 668  
 Qy 3503 attttacttgtagatt-tatttatagtatataattcgtatttaatttgggtggata 3561  
 Db 669 TAGACATGATTATATATATAAATACATATATATATGATTATATATATATGAT 728  
 Qy 3562 caatcgctttccactaaatttttaactatgattataaattttttccactcgtatat 3621  
 Db 729 AATCTAATATATAGTATTATCTCTAATATATATGTTATCTCTAATATTGGTATA 788  
 Qy 3622 ttacttataatacattttatcataatttttgaagattgagccagaacataaag 3681  
 Db 789 TAGATCTGTGAAATTAATCACTGGAGTAAGGAACCATTTTGTATAGATTATTA 848  
 Qy 3682 agacaataattctatacaagacaaatttagaaaaaattgacttttttaggttaatttaag 3741

CC fibers. They were isolated from a 23-day cotton fibre cDNA

|    |      |                                                              |      |
|----|------|--------------------------------------------------------------|------|
| Qy | 4380 | ggtactcaaaactcttggaagcctggaatgcgaagaggagaaaaaanaaaacctgcgaac | 4439 |
| Db | 307  | ACTACTCAAAACATCAGAGAACTTAAATGCACGAGGAGAAAAACAAACACCTGCACAA   | 366  |
| Qy | 4440 | agcatagagagctcacacagctcacacgaatcaaaaggacaaaaggatcaggaagaaa   | 4499 |
| Db | 367  | ATCATGAGAGATCAGACAGATCCACGCGATCGAGGAGCAGCAAGAGATCAGTAAGAAA   | 426  |
| Qy | 4500 | ctctctga                                                     | 4506 |
| Db | 427  | ATACCGA                                                      | 433  |

```

SSULT      7
T13048
ID  T13048 standard; cDNA; 1283 BP.
AC  T13048;
DE  27-MAY-1996 (first entry)
DE  Cotton fibre-specific cDNA clone E9.
KW  Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
OS  Gossypium hirsutum strain Coker 312.
PN  US5495070-A.
PD  27-FEB-1996.
PF  04-OCT-1988; 253243.
PR  04-OCT-1988; US-253243.
PR  21-NOV-1990; US-617239.
PR  18-MAY-1992; US-885970.
PA  (CETU ) AGRACETUS INC.
PI  John M;
DI  JWH; 96-130905/14.
PT  New isolated fibre-specific promoters - used for introducing
PT  altered fibre-specific characteristics into plants, partic. cotton.
PS  Example 3; Column 45-46; 48pp; English.
CC  Cotton cDNA clone E9 (T13048) was isolated from a cDNA library of
CC  cotton var. Coker 312 15-day-old boll cells using a subtractive
CC  hybridization procedure. The cDNA hybridises strongly to fibre
CC  RNA and weakly to petal RA. E9 and other fibre-specific cDNA clones
CC  (see T13034-37 and T13049-T13050) were used to screen cotton genomic
CC  libraries, leading to the isolation of genomic clones (see T13025-32
CC  and T13052-53) contg. sequences capable of promoting gene expression
CC  in fibre cells.
SO  Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

```

Query Match 4.8%; Score 265.4; DB 1; Length 1283;  
Best Local Similarity 84.3%; Pred. No. 2.9e-27;  
Matches 316: Conservative 0; Mismatches 46; Indels 13; Gaps 1;

Qy 4132 aatacaagctgtctttcttcttattgtgaatacaactgctcatagatgctgtgtcaacotttc 4191  
Db 13 ACTAATAAATCTTTGCTTTCTATTGTGAACACGCTGCTAACTTTGTCTACTTTC 72  
Qy 4192 tctcttttcaacttttactcaatgaagtgtctactagtgacggctgagcaacactgttgc 4251  
Db 73 TCTCTTTTCAACTTTTACTCACTACTGCTGTCACTAATATGCGGAGTCACACAGCTCTG 132  
Qy 4252 gcagcgctgcagctttatctgagacacaagcaacctctacagcgctcccaaatgtgct 4311  
Db 133 TCGACGGCTCGACATTTATTCTCAGACAACAACMCATCTCAGAGTGCACATTTGGCT 192  
Qy 4312 tcaaaatacgaagaacgagagctgtgaatacgaagaagcagcaatacaaaagccaaagta 4371  
Db 193 TCAAAATAGAAA-----AGCAACAGAGTGTGAATACAAACACCAAAAT 239  
Qy 4372 tccagaaagctactccaacttgagaagctgaagtccaagaagggaanaaacacaaaac 4431  
Db 240 TCCAGAAAGTACCACAAATCTGAGAGGCTAAATCTCAAGAGGAGGAAAGAACAAAAC 299  
Qy 4432 ctctcaacagctctgaagaagatcagacacagatccagatccagaagcaaaaagaagatcgc 4491  
Db 300 CTCACAACTCATGAGAGATGATCAGAGTACACGCGATGCTAGGAGAGCATGAAGATATGA 359

Db 61 TATTAATAATATTAATTAATAATTTATGGACITTTGGACTGCTGACTAATTTTCAGAA 120  
Qy 1866 tttttatttgggttttgggttttgggttttttagataattttttaaatttttgcgt 1925  
Db 121 TTTTATTGGTTTGGGTTTGGTGGTTTATGATATATTTTAAATATTCGCAT 180  
Qy 1926 aattttctgttattttgaaaggatctgaatttttttcaaaattgaaagcttaaga 1985  
Db 181 AATTTTCGTGATTTTGAAGAGATTCGAATTTTTCACAAATGAAGCTTTAGA 240  
Qy 1986 atttttactactgcasaattcogaaataagtgattttttttagaagaattaaagatt 2045  
Db 241 ATTTTACTACTGCAAAATTCAGAAATAGTGAATTTGTTTATAGAAGATTAATAGTT 300  
Qy 2046 agtattacgatttttagtttgatttgggtgaaagtaagtattgtttttgaacataattat 2105  
Db 301 AGTATTACGATTTTtagtttgatttgggtgaaagtaagtattgtttttgaacataattat 360  
Qy 2106 ttgcacataatttaagttttcttaggataaaacggaaatatcttc-ttcttttttgaase 2164  
Db 361 TTGCATATTAATTAAGTTTCTAGGAAATAAAGGAATATCTCTCTTTTGTGATAAA 420  
Qy 2165 ttactaagtcgaagacaacaacagcttttggggagcaaatatactagcttttaagtgcag 2224  
Db 421 TTACTATCTAGCAACCAACACGCTTTGGAGCAAAATTAATAGCTTTAAAGTGCAG 480  
Qy 2225 tgaactctcaaaactcgtctataactctagctgagtttgcgtgcatactagttaag 2284  
Db 481 TGTAACTCTCAAAATCTGGCTATACTCTAGGCTGAGTTTGGTGCCTACAGTAGTAAG 540  
Qy 2285 tctatagaacattacctgcacaaacgaactgagctgaggtgagatactcaaatcttct 2344  
Db 541 TCTATTAGAACTTACTGACAAAACACATGAGCTCAGGCTGAGATCTCAAACTTTTCT 600  
Qy 2345 ttttctcaattacaatattgattgattcaggttcogataataattatactacgatt 2404  
Db 601 TTTTCTCAATTAACATATGTTGATTCAAGTCCGATCTATAATATTATTATACGATT 660  
Qy 2405 atcaatttcaacttaccattatctatctattataaaataagtcagttcaactcagttt 2464  
Db 661 ATCAATTTCAATTAACCTATATACCTATTATAAATTAAGTCAGTCAATTCAGTTT 720  
Qy 2465 cgaagttcccaaaaattttgaatttttataaaatttttccctaaaaccgaattgatt 2524  
Db 721 CGAAGTCCCTAAAATTTGGAATTTTATAAATTTATCCCTAAAACCGAAATAGTAT 780  
Qy 2525 atcttcaaaatttaagtttccatttttccatcogatttcaatttccatttcttaactct 2584  
Db 781 ATCTTTCAAAATTTAGTTTCATTTTCAATCGGATTTCAATTTCACTTTTATACCTCT 840  
Qy 2585 ctattatcataattatcaataatttcaaatatttgaattatttaccatttagtccct 2644  
Db 841 CTATGATCTAATATTACATAAATTTCAAACTAATTTGAAATATATACATTTAGTCCCT 900  
Qy 2645 aagttcgaataataaatttttcaatttgaagaattatcttttccatcctaagatcaca 2704  
Db 901 AAGTTCAAACTATAAATTTTCACTTTGAAATTAATCATTTTTCACATCTAAGCATCAA 960  
Qy 2705 atttaacaaatgacacaaatttctgattagtagtagcaagtttttggctttcgaaac 2764  
Db 961 ATTATACCAAGAGCAGCAAAATTCATGATGTAGTAGTCAAGCTTTTGAAGCTTCAAAA 1020  
Qy 2765 ataaaattt-----cgaaaaaaaacaaacttaaatcattatcaatttgaacacaa 2820  
Db 1021 CATAAAAATTCAAAAAATTAACAACTTAAATCATTTATCATTTTGAACMCMAA 1080  
Qy 2821 gctttgcgaatgctgaagacttaaaagtcttcttttcttttcttttgcgaacg 2880  
Db 1081 GCTTGCCGACATCTAAGAGCTTAAAAATGCTCTTTTGTGTTCTTTTGTGCAACGG 1140  
Qy 2881 tggagagagagggaattgaagtagaccatttttttataatggttttaacataata 2940  
Db 1141 TGGAGAGAGGGAATGAAGATTGACATATTTTATTATGTTTATCATATAATA 1200

Qy 2941 ttaataaattatcataaattatacttggtaagtgcagctggggagatacgtaaagta 3000  
Db 1201 TTAATAATTTATCATATTAATCTTTTGGTGAATGTCAGCTGGGAGATAGTAAAGTA 1260  
Qy 3001 ttttaactatacttttgcgaagcttggctgtctaccaagagatcacaagttg 3060  
Db 1261 -TATAACATTTATCTTTTGCACAGATGGCTGGCTATCCAGAAGTGTCAAAGTTG 1319  
Qy 3061 agctgccttcaatgagcaattttggccataatgataaaggaatttggttagtcaa 3120  
Db 1320 AGCTGCTTCATGAGGCAATTTTGGCCATATGATAAGGCAATTTGTTAGTTCAA 1379  
Qy 3121 ctgctcagcaataattgttaaatgaaattaaatgaagtgctgtccacacacac--- 3177  
Db 1380 CTGCTCAGAGAAATGTTAAATGAAATTAATGAAGTGCCCTGTCACACACACCA 1439  
Qy 3178 aaaaaaaactatgttgggttggtaattttatacggaaatgaataattatatttta 3237  
Db 1440 AAAAAAAACAAATGTTGGTGGTGAATTTATATTACGGAATGAATGTATATTTA 1499  
Qy 3238 aataaattattgtatttagattcttaatttttggagacttccatactataatttct 3297  
Db 1500 AATTAATTTATGTTATTAGATTCTTAATTTT-GAGCATTCCTATCAATTAATCTGT 1558  
Qy 3298 a-acataattataaataatagataataaagtgtaattacttttaattacagcataat 3356  
Db 1559 ATACATAATTTAAATATAGATAATAAGATTAATTAATTTAAATACAGCATAAAT 1618  
Qy 3357 attaaattttgaactaaatttttattctatttttaatttaatttagtctatttt 3416  
Db 1619 ATTAATTTGAATCAATTAATTTTATTTCTATTATTATTTAAATTTAGTCTATT 1678  
Qy 3417 tcaaaataaatttaattcctaataaaaaaatttttcttaattgaagtgaacaaactatg 3476  
Db 1679 TCAAAATAAATTTAAATCTAAATAAAATTAATTTTCTCTAATTT----- 1725  
Qy 3477 ttatacttcaaaattataagttatttaccctgtgattttatttattagtatataa 3536  
Db 1726 -----ATTATAAATTTATTCTAACATCATATTATTACTATTATATACATAA 1773  
Qy 3537 tctgtattataattggtgggatacaactgcttccataaatttttaactatgatt 3596  
Db 1774 TTAT----- 1777  
Qy 3597 ataaatttttccaaactgttatttacttaattacataaattttatcaattttatg 3656  
Db 1778 -----AATAATTTATCATTAATTTATG 1799  
Qy 3657 gaaattgagccagcaacattagagacaacattctatacaagaacatttgaaaaa 3716  
Db 1800 GAAATGTGAGCAGAGAACATTAGAGAGCAAACTTCTATACAGACATATTAG-TAA 1858  
Qy 3717 aatgacttttagtgtaatttgaactcttcaacaaacaaacaaattcaaatcaatga 3776  
Db 1859 AATGTACTTTTAGGTAAATTTAGTACTCTTACCAACACCAAAATCAAACTAAATGA 1918  
Qy 3777 actaaataagataataacatacaggaacatttactgttaatttcaacttccataatt 3836  
Db 1919 ACCAAATAGATATATATACATACAGATATCTTACTGTATCTTACATTCCTGTATC 1978  
Qy 3837 ttattatgaaaaatacttatttatactcogaaacttactgttaatttcaacttactaa 3896  
Db 1979 ATATTATGAAGATATATATATTACTGCGCAAGTCTCTCAACAACTATATCCAA 2038  
Qy 3897 ataaagaaaac-acttaatttttaaaacttttttctatttgaagattattt 3954  
Db 2039 AAAAAATGTTGAATATATTTATTAACATTTTTTCTATATTTTGCAGATATATTT 2098  
Qy 3955 tgtattattacgttaaaattttgcataagttggaacacttcaactaatccacaca 4014  
Db 2099 TGTATATTACGTAAAAATTTGCACATAGATTGAACACCTTCTCAATCAATCCACCA 2158

Db 1973 TAGAATTTTACTACTGCAATTCAGAAAGTGAATTTGTTTGAAGAAATTAAT 2032  
Qy 2041 aagtagtattacagatTTTTgtttgttgggaagaaatgattgttttgaacata 2100  
Db 2039 AAGTAGTATATACGA TTTTGTGTTGTTGGTGAAGTATGATGTTTGAACATA 2092  
Qy 2101 attatttgacaataaagttttctagggaataacggaatactctcttttttgg 2160  
Db 2093 ATTATTTGACATAATTAAGTTTTCTAGGGAATAACGGAATATCTCTCTTTTGT 2152  
Qy 2161 aaaaattactagtcgaagacaacaaacgTTTTggggagcaatactagcttttaagtag 2220  
Db 2153 AAAATTTACTAATGACAGAACAAACACGTTTGGGAGCAATATCTAGCTTTAAGTAG 2212  
Qy 2221 tcagtgtaactctcaaaatttggtctaaactctagctaggttggtgtgtctacagtag 2280  
Db 2213 TCAGTGTACTCTCAAAATCTGTCTAATCTCTAGGCTGAGTTGGTGTCTACAGTAG 2272  
Qy 2281 taagctatagaacttacctgacaaaacgacatgactgaggtgacatcacacttt 2340  
Db 2273 TAAGTCTATGAAACCTTACCTGACAAAACGACATGACCTGAGGTGCAATCTCAACTT 2332  
Qy 2341 tcttttttctcaataaactatgcttgatcaagtttcgactatataaattttatag 2400  
Db 2333 TCCTTTTCTCAATTAACATATGTTGATCCAGTTCGCAATCATAAATTTATATAG 2392  
Qy 2401 atttatcaatttcaattacattatctatctctattataaataaagtcagttcaatag 2460  
Db 2393 ATTATCAATTCGAATACCTATATCATCTCTATTAATAATTAAGCTCAATTCAG 2452  
Qy 2461 ttttgaagttcccaaaaatttgaattttattaaattttccctaaacgcaaatag 2520  
Db 2453 TTTTGAAGTTCOCACAAATTTGAATTTTAAATTTATTCCTCAAAACGCAAAATAG 2512  
Qy 2521 ttatatcttcaaatttaagtttcaatttcaatcagatttcaatttcaatttcaat 2580  
Db 2513 TTATATCTTCAAAATTAAGTTTCATTTTCAATCCGATTTCATCTTCACTTTTATA 2572  
Qy 2581 ctctctattctataattacataaatttcaatttcaatttgaataatttcaactttag 2640  
Db 2573 CTTCTATATCTATATATACATAAATTCGAATTAATTTGAATATTTCACATTAGT 2632  
Qy 2641 ccttaagttcaaaactataaatttccatttgaagaataatcatttttcaactatagca 2700  
Db 2633 CCTAAGTTCAAAACATAAATTTTCACTTTAGAATTAATCAATTTTTCACATCAAGCA 2692  
Qy 2701 tcaaattaaccsaatgacacaaatttcagtagatttagatcaagcttttgagtttca 2760  
Db 2693 TCAAAATTACCAAAATGACAAATTTCACTATTGATAGATCAACCTTTGAGCTTCA 2752  
Qy 2761 aacataaaaattacaaaaaaacaaacttaaaatcttattcaatttgaacacaa 2820  
Db 2753 AATCAATAAATATACAAAAAACAACAACTTAAATCATTTATCAATTGACACAA 2812  
Qy 2821 gttggcagatgctgaagagcttaaaatggcttttcttttcttttgggtcaacgg 2880  
Db 2813 GCTTGGCGATGCTAAGAGCTTAAAGTGCCTCTTCTTTCTTTTGTGCAACAGG 2872  
Qy 2881 tggagagagagggaaatgagagtgaccataatttttattatttttaacataaata 2940  
Db 2873 TGGAGAGAGAGGGAATGAGATTGACCATATTTTTTATTATGTTTTAACAATATA 2932  
Qy 2941 ttaataatttatactatacttttgtagaatgtgacgtggggagatagcTaaagta 3000  
Db 2933 TTAAATATTTAATCATATATATACCTTGGTGAATGTGACATGGGGAATAGCTAAGTA 2992  
Qy 3001 ttttaacattatactttgcaagcgttgctgtgtctacccagagtgatcaagttg 3060  
Db 2993 TTTTACATATATCTTCTTGACAGCACTTGGTGTCTACCCAGATGATCAAGTTTG 3052  
Qy 3061 agcttcctcaagagccaaatttttgcataatgataaggaactttgttagtcaa 3120

Db 3053 AGCTGCTTCAATGAGCAATTTTGGCCATATGATAAGCAATTTGTTAGTCAA 3112  
Qy 3121 ctgtctacagataatgttaaaatgaatataaataaggttgccctgttcacacacaaa 3180  
Db 3113 CTGCTACAGAAATAGTAAATGAATTAATAAAGGTGGCTGGTCAACACAA 3172  
Qy 3181 aaaaaactaagtttggttggttatttatattacagaaatataattattttaa 3240  
Db 3173 AAAAATCAATGTTGGTGGTCAATTTTATATACGGAAGTAAATATATATTTAAA 3232  
Qy 3241 taaaattattgtatttagcttcaatttttgagcattccactataaattcgttaac 3300  
Db 3233 TAAATATTGTATTAGATCTTAATTTTGGAGCATTCATATAATTTCTGAAC 3292  
Qy 3301 ataatattaaaatagtaataaaggttaatttaacttttaattcaagcattatata 3360  
Db 3293 ATAAATTAATAATATAGTAAATAAAGGTAAATTAACTTTAAATACAGCATATA 3352  
Qy 3361 aattttgaaactaaattttttttcttatttttatttaatttagtctatttttcaa 3420  
Db 3353 AATTTCGAATCAATTAATTTTATTTCTAATTAATTAATTTAGTCTATTTTCAA 3412  
Qy 3421 aataaaatttaacttaataaaaataatttttcttaatttggaacaactcgtttat 3480  
Db 3413 AATAAAATTAATTAATAAATAAATTTTCTTAATGTGAACACATCATGTTAT 3472  
Qy 3481 acttcaaaatttaagatttatatttacccttgatgatttatttagtattatattct 3540  
Db 3473 ACTTCAAAATTAAGTATATATATTACCTGATGATTATTATTAGTATATTAATCT 3532  
Qy 3541 gattataattatgtgggacacatgctttccactaaaattttaactatgattataa 3600  
Db 3533 GATTATTAATTTGGTGGGACATACGCTTCCATTAATTTTACATGATTATTATA 3592  
Qy 3601 atttttttcaactcgtattatttacttataatacaattttatcattttttagtaa 3660  
Db 3593 ATTTATTTCACATGATATTTTACTTATTAATCAATTAATTTATCATATTTAGTAA 3652  
Qy 3661 ttgagcaagaagaaacttaaggaacaaattctaaacaaagcaatttagaanaaatg 3720  
Db 3653 TTGACACAGAAACATTAAGACAACTTCTAAACAAAGCAATTTGAAAAAATG 3712  
Qy 3721 tcttttagtgatttttaagactcttaacaaacaaaaattcaaatcaaatgaacta 3780  
Db 3713 TACTTTTAGTAAATTTAAGTACTCTTAACAAACAAAATTCAAATCAATGAACTA 3772  
Qy 3781 aataagaataataaactacggaacatttacttgatttaccatcccaattttat 3840  
Db 3773 AATAGATATATAAATACGAGACATCTTACTCTGATATTTACATGTCATATTTAT 3832  
Qy 3841 tatgaanaataacttatactacagactaaagttgtgcacaaatttatctaaata 3900  
Db 3833 TATGAAAATAATCTTATATATACGACATAAATGTTGTCACAAATTTATCAATAA 3892  
Qy 3901 agaaaacacttaatttttataacattttttcatatttgaagaatttatatttgata 3960  
Db 3893 AGAAAACACTTAATTTTATAACATTTTTCATATATTAGAGATATATATTGTATA 3952  
Qy 3961 tttacgaaaaattttgacatgattgagcccttttaacataatccacacataagtc 4020  
Db 3953 TTTACGTAAATATTGTGACATAGTGAAGACCTTTCTTACATATCCACACATAGTC 4012  
Qy 4021 aagtagtagatgagaatttgtaacaaacagctggggcacaattcccaacacattct 4080  
Db 4013 AAGTATGTAGATGAGAAATGGTACAAACAGCTGGGGCCAAATCCACAAACATCTC 4072  
Qy 4081 tcatctctctataaaaagctgtgtacacatagacacaaatccacacaaataacgct 4140  
Db 4073 TCATCTCTCTCTATAAAGGCTGTACACATAGACACAAATCCACACAAATACAGT 4132  
Qy 4141 tctttttttctatttgattaaactgggtgatagcttgctgcaccttttctcttttc 4200  
Db 4133 TCTTTCTTCTTATTTGATTAACATGGCTATAGCATGCTGACCTTCTTCTCTTTC 4192

Db 3841 TATGAAAAATACTTATATTACTGCACTAAATGTCACAAATATTATCTAAATAA 3900

Qy 3901 agaaaaacctaatttttataaacttttttcatatttgagaagattatttgata 3960

Db 3901 AGAAAAACCTTAATTTTATACATTTTTTCATATTTTGAAGAATATATTTGTATA 3960

Qy 3961 tttaagttaaaattattgacatagattgagcaacttttcaacaataccacaaatagtc 4020

Db 3961 TTAAGTAAAAATTATGACATAGATTGACACTCTCTAACATAATCCACCAATAAGTC 4020

Qy 4021 aagtatgatagagaatttgtaacaaacaagctggggccaaatccacacaaacatctc 4080

Db 4021 AAGTATGATAGAGAAATGGTACAAACACGTGGGGCCAAATCCCAACAAACATCTC 4080

Qy 4081 tcatctctctataaaagcttggtacacatagacaacaatccacacaaatacaagt 4140

Db 4081 TCATCTCTCTATAAAAGCTTGCTACATAGACACAACATCCACACAAATACAGT 4140

Qy 4141 tcttttcttttattgataaaccatgctcatagacttgctacacttctctctttc 4200

Db 4141 TCTTCTCTCTATTAATTAACCATGCTCATAGACTGCTCACCTCTCTCTCTTTC 4200

Qy 4201 caacttttactaaagtgtctcaactagtgacggtagccaactgtttggagagcgt 4260

Db 4201 CAACTTTACTCATAGTGTCTCACTAGTACCGTAGCCCACTGTTTGGCAGGGCT 4260

Qy 4261 cgagctttatcgagacacagcaactctcagagctccacacattggtctcaaaatc 4320

Db 4261 CGAGCTTTATTCGAGACAACGCACTCTCATGAGCTCCCAATTTGGCTCTCAAAATC 4320

Qy 4321 gaaagcagcagagcttgtaatacgaagaacgaatacaaacagccaagtatacagaga 4380

Db 4321 GAAAGCAGAGAGCTTGTAATACGAAGACGAGATCAACAAACGCAAAATATCAAGAGA 4380

Qy 4381 gtactcaaaacttgagaagcctgaagtgaagaaggaaaaaaaacccctgaaac 4440

Db 4381 GTACTCAAACTTGAGAAGCCTGAAGTCAAGAGGAAAAAAGAAAAAACCCTGCAACA 4440

Qy 4441 gcatgaagagatacacaagctcacagcaatacaaggacaaaagagatcagaagaana 4500

Db 4441 GCATGAAGAGTACCAAGAGTCAACAGAAATCAAGAGCAAAAAAGATCAGGAAGAAAA 4500

Qy 4501 tctcagagaattcccccggcgttcagcagctagcagaattctcggccgcgtcagact 4560

Db 4501 TCTCGAGAAATTCOCOCGGCGCTCAGCGCTACGAGATCTTCGGCCCGCTCAGGCTT 4560

Qy 4561 gaactatgatcaactggtgcaatgccaatcatcagcgaatttctcgttatctgttaa 4620

Db 4561 GAATCATATGACACTGGTGCATGTGCCATCATCAGCAGTAATTTCTATGTATATCTGA 4620

Qy 4621 tatatgttaataaaaaagattggtgattggaattggtggtgcatctctccatgaact 4680

Db 4621 TATATGTATTATAAAAAAGATGGTGATGGGAATGTGTGTGTGCATCTCTCCAGCACT 4680

Qy 4681 aatgtagaactcttgcatcacaagaattctaaatggttatagttatgttatagtt 4740

Db 4681 AATGGTAATCTCTTTCGACACATAGAAATCTAATGTTTATAGTTATATAGTGT 4740

Qy 4741 agttgtatggaatttaatttaaatggtgtatctaaatgtaactcaactggtcatt 4800

Db 4741 AGTTGTATGGAATTAATTTAATGTGTATCTAAGTGTAAACATCACTGGCTGAT 4800

Qy 4801 tatgtatgtatgtatttacttaataatgataatgcaatgtattgtaataaactgct 4860

Db 4801 TATGTATGTATGTATTACTTATATATATGCATATGTATTTAATTAACATGCT 4860

Qy 4861 tgatcatataactctctactataataataagcagactttgttttaacttttacc 4920

Db 4861 TGATCATATACTCTCTACTATTAATAAATAGCACTGTTTGTTTAAMCTTTTAC 4920

Qy 4921 aagttaagcagataaataatgacaaataataacaggttttaagtcaatgtatgc 4980

Db 4921 AAGTTAAGCATGTATAAATATGACAAATATAATACAGGTTTATGTTCAATGTATCT 4980

Qy 4981 actcttagtagttttgatgacttaataacatttaacaaattccatttaaaatttaa 5040

Db 4981 ACTCTAGTAGTTTATGATGACTTAATTAATTAACAAATTCCCTAAAATTTTAA 5040

Qy 5041 taataataaacaataatattgtaataataacatttaaggaacaaataaataa 5100

Db 5041 TAAATATAACAAATATTATGTAATATATACATTTAAAGCACAAAAAATGAATTA 5100

Qy 5101 ataaaaataaataaataatgttataatattgtaataataatgtacatattctta 5160

Db 5101 ATAAATATAATACAAATAATTGTTATATATGTAATATATGTAATATGTAATATCTTA 5160

Qy 5161 actgaataagggttaactatactactccataaatttgatttaaatattttatacctac 5220

Db 5161 ACTGAATAGGGTTAACTATAATCTCTAAATTCAGTTTAAATATTTTATACCTAC 5220

Qy 5221 catattattagactctttttataataataaatttttaataacatttaataaac 5280

Db 5221 CATATTATTAGACTCTTTTAAATATTAATAATTTAATTATACAAATTAATTAAC 5280

Qy 5281 tattattatcttaactaaatactaaatttttataacattataataatcttaata 5340

Db 5281 TATTATTATCTTAACATAAACTAAATTTTATTAACCTATTATTAATTAATTAATTA 5340

Qy 5341 tcttatcaatttaaaactcttaatacttaataaatttaataatcttaatactta 5400

Db 5341 TCTTACTAATTTAAACTCTTAATTTCTTAATTTAATTTAATTTAATTTAATTTAAT 5400

Qy 5401 ttgtcaactctccaccagcagtagatgctggaccgaatccggagattacatccggcat 5460

Db 5401 TTGTACTCTCTCCACCAGCAGTAGTCTGGACCGAATCCGGAGATTACATCCGGCAT 5460

Qy 5461 tgagatggcgtctatcgggtttggcgcgcgctgccaattgcgcctatagtagtgct 5520

Db 5461 TGAGATGGCGTATCAGGTTTGGCGCGCGGTACCAATTCGCCTATAGTAGTGTCT 5520

Qy 5521 attacgcgcgtcactgcgtcgggttt 5547

Db 5521 ATTACGCGCGTCACTGCGTGGGTTT 5547

## RESULT 2

T73870

ID T73870 standard; DNA; 5518 Bp.

AC T73870;

DT 26-JAN-1998 (first entry)

DE Cotton fibre promoter clone 4-4(6) construct, pCGN5610 (Version II).

KW promoter; fibre-specific; transcriptional factor; promoter;

KW altered phenotype; colour; melanin; indigo; ss.

OS *Gossypium hirsutum* cv. coker 130.

PN W09640924-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; D09897.

PR 07-JUN-1995; US-480178.

PR 01-JUL-1996; ZA-005572.

PA (CALJ) CALGENE INC.

PI McBride K, Pear JR, Perez-Grau L, Stalker DM;

DR WPI; 97-052125/05.

PT DNA construct contg. gene of interest controlled by cotton fibre

PT transcriptional factor - used to produce altered phenotype cotton

PT fibre cells expressing genes affecting pigmentation

PS Example 5; Fig 3A-J; 95pp; English.

CC The present sequence is a 4-4 cotton fibre expression cassette (version

CC II) from promoter construct pCGN5610. The lambda genomic phage clone used

CC to form this construct was designated 4-4(6). DNA constructs containing

CC cotton fibre-specific transcriptional factor promoters are useful to

CC produce cotton fibre cells with altered phenotype, especially altered

CC colour. Genes involved in the production of melanin (e.g. tyrosinase

CC gene and ORF438 encoded protein from *Streptomyces antibioticus*) and

CC indigo (mono-oxygenase genes possibly in conjunction with a

CC tryptophanase gene) are of interest. The promoters of the invention are

CC reliable and permit expression of a protein selectively in cotton fibre



PF 07-JUN-1996; U08997.  
PR 07-JUN-1995; US-480178.  
PI 01-JUL-1996; ZA-005572.  
PA (CALJ) CALGENE INC.  
FI McBride K, Pear JR, Perez-Grau L, Stalker DM;  
DR WPI; 97-052325/05.  
DR P-PSDB; W21899.  
PT DNA construct contg. gene of interest controlled by cotton fibre  
PT transcriptional factor - used to produce altered phenotype cotton  
PT fibre cells expressing genes affecting pigmentation  
PS Claim 22; Fig 2A-J; 95pp; English.  
CC The present sequence is a 4-4 cotton fibre expression cassette (version  
CC 1) from promoter construct pCGN5606. The lambda genomic phage clone used  
CC to form this construct was designated 4-4(6). DNA constructs containing  
CC cotton fibre-specific transcriptional factor promoters are useful to  
CC produce cotton fibre cells with altered phenotype, especially altered  
CC colour. Genes involved in the production of melanin (e.g. tyrosinase  
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and  
CC indigo (mono-oxygenase genes possibly in conjunction with a  
CC tryptophanase gene) are of interest. The promoters of the invention are  
CC reliable and permit expression of a protein selectively in cotton fibre  
CC to affect qualities such as fibre strength, length, colour and dyability  
CC as required. The construct and methods can also be used for the  
CC introduction of other advantageous genes into a cotton plant, e.g. a  
CC plant hormone. In particular, fibres from a plant producing coloured  
CC fibres may be used to produce yarns and/or fabrics that do not require  
CC dyeing.  
SQ Sequence 5547 BP; 1889 A; 808 C; 822 G; 2028 T;

Query Match 100.0%; Score 5547; DB 1; Length 5547;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5547; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 actaaagggaacaaaagctggagctccaacgcgggtggcgccgctctagaacatgtgat 60  
Db 1 ACTAAGGGAACAAAAGCTGGAGCTCCACGCGGTGGCGGCCGCTCTAGAACATGTGGAT 60  
Qy 61 cccccgttgactaaacaaacatgggaagatttgcgtgtaaaaaataaaagaagcttact 120  
Db 61 CCCCCGTGACTAAACAAACATGGGAAGATTTCGTGTAATAAAATAAAGAGCTTACT 120  
Qy 121 caataacacttcttgtaattgtatatacaaaagctcaatgaaacataactcaatacact 180  
Db 121 CAATAACACTTCTTGTAATTGTATACAAAAGCTCAATGAAAAACAATACTCAATCACT 180  
Qy 181 ttttttcagctatttaactctcttatactagctgaactacacaaacttagctaaanaa 240  
Db 181 TTTTTCAGCTATTTAACCTCTTATATAGCTGAACATACCAACATTTAGCTAAAAA 240  
Qy 241 ataggataacctaataagaaaactcaactcagatataaacaatgattttagctaacat 300  
Db 241 ATAGGATACCTAATAAGAAAACCAACTCAGATATAAACATGATTTTAGCTAACAT 300  
Qy 301 ttaacaacttattgaacactaattgaaatttcaactctgcgtatagcccaagatttag 360  
Db 301 TTAACAACCTTATTGAACATTAATTGAATATTCTCTGCTGATATGCCCAAGATTTAG 360  
Qy 361 gccactaacagatttgggtgaaatttaacactgcatgcatattgtaactggttgaaca 420  
Db 361 GCCACTAACAGATTGGGTGAACTTTAAACATGCAATGCAATTGTGAATGTTGAACA 420  
Qy 421 agttttttgcaattatttactatatagaactgtttgattgagttgagtaacacagagct 480  
Db 421 AGTTTTTGCATTATTACTATATGAACGTGTTGATTAGTGTGAGTTACACATGAGCT 480  
Qy 481 tgaagctcaactcaaaattttctcaattctaaaggtgacgaacaaactgaagccgggg 540  
Db 481 TGAAGCTCACTCAAAATTTTCTAATTCTGAAGGTGATCAGCAACTTGAAGCCGGGG 540  
Qy 541 cgtacagagctcgagattgatttttagtttaataaagaagcattatgttttaaacct 600  
Db 541 CGTACGAGCTCGAGATTGATTCTGATTGTAATAAAGACGATTATGTTTAAACT 600

Qy 601 attatggactttttggactatgaactgttgaggactttattttttttttattgtctt 660  
Db 601 ATTATGGACTTTTGGACTATGAACGTGTGGAGCTTATTATTGTTTTTATTGCTT 660  
Qy 661 tttttgattgaaataattatttttaaaactgcaaaattatagtttttacaactaag 720  
Db 661 TTTTGGATTGTAATATTATTTTAACTGCAAAATTATAGTTTTCACAACTAG 720  
Qy 721 tccagattttcaaaactccataacttagaattttgcgtcaaaaataagaaatcatta 780  
Db 721 TCACAGTTTCAAAATCCATAACTTAGAATTTTGCCTCAAAATAAGTATCATTTA 780  
Qy 781 agtgatttttctgataaaaaataaataatttttaacagagattttctctaaaattgg 840  
Db 781 AGTGTTTTCTGTAATAAAATAAATAAATTTTAAACAGATTTCCTCAAAATGG 840  
Qy 841 aaattgatttaccaaaatttagtatgtcaaaacacatgtttatgttacaggcgatc 900  
Db 841 AAATGATTTACCAAAATTAGTATGTCAAAACACAGTTTATATGTACAGGCGATC 900  
Qy 901 gctcaggaacaaacatctaggcgggtttgaggtttacagagcgagtggttcattt 960  
Db 901 GCTCAGGAACAAACATCTAGCGGGTTTGGAGTTTACAGGCGAGTGCGGCTCATT 960  
Qy 961 gaggtaagttagtagggcgagtttagatgtgcattacaagtcagaagatttgtaaa 1020  
Db 961 GAGTAGATTAGTAGGGCGAGTTTGGATTGCAATTCAGCTCAAGATTGTGAAA 1020  
Qy 1021 ctgcagaaatgattgtatgattgtgcagattcaacaaatattttttctttgtgtg 1080  
Db 1021 CTCGATGAATGATATGATGATTGTCTCGATTACGAATATGTTTTTCTTTGTGTG 1080  
Qy 1081 tgtttatctcgtgtgataatataatgatttttttaccatttatggcagtgac 1140  
Db 1081 TGTTTATCTCTGCTGATAGATATATGATTGTTTATCCAACTCTTAGCGATGTAC 1140  
Qy 1141 attggtgctatttctaattgaattgattgtattatgaacatgtatgcattcttcac 1200  
Db 1141 ATTGTGGCTATTCAAATTAATGATTGTGTTATATGAACATGATGCTCTGTCTAC 1200  
Qy 1201 aaagcatggaactcagctcactcttctgtttaaagatacagatgcaagtttaacatg 1260  
Db 1201 AAAGCATGGAATCTAGCTACTGCTTCTGTGTAAGATACGATTGCAAGTTTACATC 1260  
Qy 1261 cttaactattttgattgtcttgcagctatgtccactacagcgggttggaatgat 1320  
Db 1261 CTACTATTGTTGATTGTCTGCTGATGCTACATTCAGTGGGTTGGAGATAT 1320  
Qy 1321 ggtaagggaagattttgacagtttaagatttgactatcgtgtgttaaccacat 1380  
Db 1321 GGTAGGAGGAAGTTTGACAGTTTAAAGATTGCACTATCTGTGTTTAAACCATAT 1380  
Qy 1381 ttgtatggcatttgcactggttattggtgctgcagccacatctctgttggaaat 1440  
Db 1381 TTGTATGGCACTTGCAGCGGTTATGGTGGCTGAGCCCAATCTGTTCTGGAAT 1440  
Qy 1441 ttatctgtcactgttggtgacttgcataattttgtgtgtgttttggatggagc 1500  
Db 1441 TTATCTGTGACTGTGGGCAATGCTACAAATTATTTGTTGGTGTGTTGGATGGAGC 1500  
Qy 1501 gtcgagggaactctattgtgtgtgtgaggtgggtgaagaaatttcyaaaaaaatt 1560  
Db 1501 GTCGAGGGAACCTCTATTGTTGTTGCCGAGTTGGTAGGAATTTGCAAAAAAAT 1560  
Qy 1561 tgaattgtgttttgcataaaattggaataacataatgaatgattcgaatttggc 1620  
Db 1561 TGCATTGTGTTTTCTCAAAATATTCATATACATATCATTCATTCATTTGGCT 1620  
Qy 1621 aattgaagatataaaattctctatgatactgctctgttttaacattatgtgttt 1680  
Db 1621 AATTGAAGTATAAAATCTCTATGATACCTGATCTGTTATTACATTATGTGTTT 1680



Qy 3232 attttaaaataaaaaattgtttattgattgttataatttttggcgaccttccataataaa 3231

Db 17044 ATATATATATATATATATATATACAAATTAATTAATTCAGATTTAGTGATTAATAAATAT 16985

Qy 3292 tttcgtacaataattttaaaatactgataataaagtgttaattaaactttaattacaag 3351

Db 16984 TTTTATTACTATATATATTTAATTGAAATTAATAATATGATATATATATATATAAT 16925

Qy 3352 ataataattaaaatttgcatactaaattttttttcttataattttaataatttagtcta 3411

Db 16924 TGAATTTGAATTTTTTAAAAAAATCAATTTTAAATTTTTATATATAAAAATTTTTTAT 16865

Qy 3412 tttttttcassataaaattttaaattcassataaaaaaattttttcttcattgttgaacac 3471

Db 16864 AATTAATTTAAATATTTTTTATTATAAAATATTATATAT-----AAAT 16819

Qy 3472 tcatgtctactctcassatttaagtatctatttatcttccttgatgatttattttagtat 3531

Db 16818 AGTTATTAGTATATTTTATATAAATCAATTTTTTTTTTAAAAAAAATATTTTTTAGTG 16759

Qy 3532 attaatcttgattataattttatgtgtgggatacaatcgctttccacaataattttaactat 3591

Db 16758 TTAATTATACATAAATTTATGATAGGGGGAATAATTTATTTTCATTTTTTATAT 16699

Qy 3592 -gatttataaaatttttcttacaatcgtattattacttatatacataaatttatcataat 3650

Db 16698 ATATATATATATATAATTAATTTATTCAGATTTAGTGATTAATAAATATTTTATTAT 16639

Qy 3651 tttatggaaattgagccagaacaattacagagacaaattctatacaagaacattta 3710

Db 16638 ATTATATATATTA-----ATGAAAAATAAATATATATATATATATATA 16590

Qy 3711 gaaaaaaatgacttttttagtgatttttagtactcttatacaaaaacaaaatctcaact 3770

Db 16589 TAAATATAAATGAATTTTTAAAAATATTTTAAATTTTTTATATATAAATTAATTC 16530

Qy 3771 aaatgactacaataagaataatacaatacagacacactctacttgtaattctacattcc 3830

Db 16529 TTTATTATATTTTTAAATTAATTTTATTATAAATAATTTATATAAA-----AATAGTTA 16473

Qy 3831 ataattttttatgaaaaatactctatactatctgcagacaatggtgtcacaataatt 3890

Db 16472 TTAGATATAAATTAATAAATATTTTTTTTTTAAAAAAAATTTTTTTAGTTTTAAT 16413

Qy 3891 atctcaataaagaaaaaacacttaatttttatacaattttttataattttgaagattat 3950

Db 16412 ATATA-----TAAATTTATGATAGGGGGAATAAATTTATTTTCATTTATATATA 16359

Qy 3951 attttgtatatttaagtaaaaaatttgacatactgtgcagacctcttaacatactcc 4010

Db 16358 TATTATATATATATATACATAAATTAATTCAGATTTAGGTATAAATTAATATTTT 16299

Qy 4011 accataagtcagatgtagatgagaagaattggtacaacaacagtgggggccaattccacc 4070

Db 16298 ATTATACTATATATATTAATTTGAAATTAATAATATGATATATATAAATATATGA 16239

Qy 4071 aaacactctctattctctctctataaaagggtgctacacatagacaacaatccacac 4130

Db 16238 ATTGAAATTTTTTAAAAAAATCAATTTTAAATTTTTTATATATAAAAAATTTTTTAAAT 16179

Qy 4131 aaatacagctgtttttttttcttatttgatttaacatggctgatagc-----tctgcacc 4187

Db 16178 TATTTAAATTAATTTTATATATAAATTAATTTAATAAATAATGATTATTAAGTATAT 16119

Qy 4188 ttcttctcttttccaaattttctctcaatgagtctcactagtagccggtagccagctgt 4247

Db 16118 TTAATTAATCAATTTTTTTTTAAAAAAAATATTTTTTAGTTTTTAATATACAAATAAT 16059

Qy 4248 ttggcga-----gcgctgcagctttatttcgagacaagcaacctctcagagctccacc 4303

Db 16058 TTTATGATAGGGGGAATAAATTTATTTTCATTTTTTATATATATATATATATATAT 15999

[illegible]

|           |                                                                                                                                                                   |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | Pterygota; Neoptera; Endopterygota: Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.                                                     |
| REFERENCE | 1 (bases 12511 to 12682)                                                                                                                                          |
| AUTHORS   | Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.                                                                                          |
| TITLE     | Drosophila mitochondrial DNA: a novel gene order                                                                                                                  |
| JOURNAL   | Nucleic Acids Res. 10 (21), 6619-6637 (1982)                                                                                                                      |
| MEDLINE   | 83090428                                                                                                                                                          |
| REFERENCE | 2 (bases 5269 to 5695)                                                                                                                                            |
| AUTHORS   | Clary,D.O., Wahlleitner,J.A. and Wolstenholme,D.R.                                                                                                                |
| TITLE     | Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes                           |
| JOURNAL   | Nucleic Acids Res. 11 (8), 2411-2425 (1983)                                                                                                                       |
| MEDLINE   | 83220794                                                                                                                                                          |
| AUTHORS   | 3 (bases 404 to 5272)                                                                                                                                             |
| TITLE     | de Bruijn,M.H.                                                                                                                                                    |
| TITLE     | Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code                                                                                  |
| JOURNAL   | Nature 304 (5923), 234-241 (1983)                                                                                                                                 |
| MEDLINE   | 83245048                                                                                                                                                          |
| REFERENCE | 4 (bases 804 to 1778)                                                                                                                                             |
| AUTHORS   | Satta,Y., Ishiba,H. and Chigusa,S.I.                                                                                                                              |
| TITLE     | Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species                                                     |
| JOURNAL   | Mol. Biol. Evol. 4 (5), 638-650 (1987)                                                                                                                            |
| MEDLINE   | 88174373                                                                                                                                                          |
| REFERENCE | 5 (bases 5268 to 13619)                                                                                                                                           |
| AUTHORS   | Garesse,R.                                                                                                                                                        |
| TITLE     | Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations                                                                      |
| JOURNAL   | Genetics 118 (4), 649-663 (1988)                                                                                                                                  |
| MEDLINE   | 88212147                                                                                                                                                          |
| REFERENCE | 6 (bases 441 to 2967)                                                                                                                                             |
| AUTHORS   | Satta,Y. and Takahata,N.                                                                                                                                          |
| TITLE     | Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup                                                                            |
| JOURNAL   | Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)                                                                                                           |
| MEDLINE   | 91088557                                                                                                                                                          |
| REFERENCE | 7 (bases 14215 to 14512)                                                                                                                                          |
| AUTHORS   | Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.                                                                                  |
| TITLE     | Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods                                                                              |
| JOURNAL   | Science 258 (5086), 1345-1348 (1992)                                                                                                                              |
| MEDLINE   | 93088057                                                                                                                                                          |
| REFERENCE | 8 (bases 14917 to 19517)                                                                                                                                          |
| AUTHORS   | Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.                                                                                                              |
| TITLE     | Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA                                                              |
| JOURNAL   | Mol. Biol. Evol. 11 (3), 523-538 (1994)                                                                                                                           |
| MEDLINE   | 94285822                                                                                                                                                          |
| REFERENCE | 9 (bases 1 to 408; 13319 to 19517)                                                                                                                                |
| AUTHORS   | Lewis,D.L., Farr,C.L. and Kaguni,L.S.                                                                                                                             |
| TITLE     | Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons                                                     |
| JOURNAL   | Insect Mol. Biol. 4 (4), 263-278 (1995)                                                                                                                           |
| MEDLINE   | 96423163                                                                                                                                                          |
| REFERENCE | 10 (bases 1 to 19517)                                                                                                                                             |
| AUTHORS   | Lewis,D.L., Farr,C.L. and Kaguni,L.S.                                                                                                                             |
| TITLE     | Direct Submission                                                                                                                                                 |
| JOURNAL   | Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA                                    |
| FEATURES  | Location/Qualifiers                                                                                                                                               |
| source    | 1..19517                                                                                                                                                          |
|           | /organism="Drosophila melanogaster"                                                                                                                               |
|           | /organelle="mitochondrion"                                                                                                                                        |
|           | /db_xref="taxon:7227"                                                                                                                                             |
|           | /note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains" |
| tRNA      | 1..65                                                                                                                                                             |
|           | /gene="mt.tND5"                                                                                                                                                   |

```

RESULT 15
DMSI37541/c
LOCUS DMSI37541 19517 bp DNA circular INV 04-APR-2000
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

VERSION AL031746.9 GI:6594243  
 KEYWORDS Htg.  
 SOURCE malaria parasite *P. falciparum*.  
 ORGANISM Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 REFERENCE 1 (bases 1 to 67970)  
 AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrett, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.  
 For more information about this sequence or the Malaria Project, see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum). IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from *E.coli*, yeast, vector, phage etc.

FEATURES  
 source Location/Qualifiers  
 1..67970  
 /organism="Plasmodium falciparum"  
 /strain="3D7"  
 /db\_xref="taxon:5833"  
 /chromosome="1"  
 gene complement(1748..3276)  
 /gene="MAL1P3.01"  
 CDS complement(join(1748..2598,2748..2848,2990..3276))  
 /gene="MAL1P3.01"  
 /note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPP0006 family eg to YB055C/YB0512/YB0511, YBFS\_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)"  
 /codon\_start=1  
 /product="conserved hypothetical protein, UPP0006 family"  
 /protein\_id="CAB63556.1"  
 /db\_xref="GI:6594244"  
 /translation="MKLVFHYIKINLVFYISIIPLKSNLSKILNDLYRISTVKNYKVLQIKRSLKLNHNIRKMDNSSFIDGSLNDGPGVYNSKKNENDGLWLNRAKNNVNDKIITPCTCLAEIDSKLKICETYPDSEKFLYLSAGVPIPCFEDKNNKEEKEIIAAKYEDEFIKYFKNEQVNSKMGNNKKICGDKMDNNWELLKXLDITPEFYKNEKDEYLENLAKNKIKIPNRYVICGEIGLDFDLRYCSKTIQIKPIQLKLVQVNP LPMFLHNGCSETFFKIVDIYKFLPKNGVIRSPDKEDIVHIVQNKMLYIGVNGCSLKSLENINAVKIKPLMLLILEDPACWGCVKTKYASETYIKDTYKATYNLAKIKNIICDDWTFKERNPEYPIA"  
 misc\_feature complement(2599..2610)  
 /gene="MAL1P3.01"  
 /note="potential splice acceptor sequence"  
 misc\_feature complement(2742..2747)  
 /gene="MAL1P3.01"  
 /note="potential splice donor sequence, atg/gttaa"  
 misc\_feature complement(2849..2861)  
 /gene="MAL1P3.01"  
 /note="potential splice acceptor sequence"  
 misc\_feature complement(2884..2989)  
 /gene="MAL1P3.01"  
 /note="potential splice donor sequence, aag/gtaaaa"  
 gene 5005..5496  
 /gene="MAL1P3.02"  
 CDS 5005..5496  
 /gene="MAL1P3.02"  
 /note="MAL1P3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"  
 /codon\_start=1  
 /product="hypothetical protein, MAL1P3.02"  
 /protein\_id="CAB63557.1"  
 /db\_xref="GI:6594245"  
 /translation="MKLLNRFVILCPILPILFVSLVGLWNNRNINHFETENAAKAMKRLSGEISGILKIDGDELKILKNDKSTGWSYFSINLEEKXYSDTLFRRKKQZINEAWYIKIDRQBFYILNNDENETARFVLENNDFITQSGFKLIDITGS

misc\_feature 8020..10389  
 /note="possible cent. region of very high [A+T] content"  
 gene 14884..20352  
 /gene="MAL1P3.03"  
 CDS 14884..20352  
 /gene="MAL1P3.03"  
 /note="MAL1P3.03, putative ABC transporter, len: 1822 aa"  
 /codon\_start=1  
 /product="putative ABC transporter"  
 /protein\_id="CAB63558.1"  
 /db\_xref="GI:6594246"  
 /translation="MTYKRWGISKNGKKNKSKQNSIYPLNFSLDFWIRPLNDLIKGDIGELPNICRNFDPVPTASKLEENLRIEDVDSFSEKSNSEWLEHNSQDASEKKVYNINHNILMSLKATPFRILITISFYILETILVTLGGKFIPTYRLEGQKIPVYISLADPQVYFSLGVVIMFPHFLPFPALLFPYHLPTINLAKVMTPLTKINLCSNNHLQNPDPATYNTYKPSQTEIDELSDPLSIGKNASSSSGCIKNNKNIDNNKFVENDTIIINFIKSTKMKKSDLSNENKPLNWNYNIMFDVSPVSPYFVTSICMLNFPVKPIMSPYFVRIKIGSNVGIATWLSIALYSAMILPEFLPLSKYKTLIDYDRKIDNMHVLKEFLIDMPWNSFAPKYINLFRMKKNCKYKRLKLSNIGVFSISSIDVVEVFIPLYADLRKKEEIKFTSILMPLYKTLISNVNPNLVNVMVNGVYNKRNQYINDHLYTDINDINTFMYRTNEDYINVDKTFIQWNTISDDGTSNHLKKNWIKNLIKMYTFPFYKMYHKNINQKILSGKLNDVDDTNKIKQFQRHSNNTSTNTNSSHIEKKEETENHSSNSTMSNEPKKKNNNDNIIKLENCSGFLSTDKDCDDHDLKANIKNFERNLSLAIIGVSGSKAPFSSILCDPNTMGNLYTENFAPKMPILVQPNWLPFGNTRSMILPGNEYNLYKTYTLQSELNDLSTIEHGDMYINDBNLSGQKVRICIALRALTVEIHMKLCTDYEKKLIQNEILKDLINNNISYNNKSKLVNTPFPNENYQLCQADDNDNLYLDLIDPISLDPSSIKKPLCKEDKNSIFKDCSFISIMKSSTLDMFLIEDLDVQVYENIPELDQTKLYKGNISEMEKNINLITESHGSGVSNLNTIDYRILPLDEVLHVRHNSKNMRYEATFVNGTNSVSEFSDISNKEYIKMKKKRYKKEHNNKNNKNNNNNNKDDHININNDNHNRYDILGLPSTDDSPVSSLSGNE"  
 TLDYTTNSNNSSKEIVLPKIDTHEEPNKSMPFPYSSNMKNPNKPNYEDNSSFKGSISETLYLWFOQVQVGLLSVPLMLISIFDIEIKYVPLNMSISNNKRESDYQLQVRYLEYFVLPIISLIVTSCFCFMSLIGYNTISKVNHNILSILNAPLIFITYNNGLKINRILDISADPQFKRIKRYAFPIFRCILSSLIITMROCIPIFPFVILIFFVFRFSGCKEQRILYSCPLCNITNALSGLNINITYKNTYLDYDTEHIFNFRISYFKNLNLWASLYIKPILLLITTYIHMPFLASGIIKLYKKNVRIESTLYCISFARLGVITFLCDDYILKHECCQVBLEFASIKSNENASMKRENELNWTITQYKKNENISDKISAIVYKNSLSINSQSDSSKKYKIFENYVTSYKIPILVNGTQYIIDEPSKINIMYALKKQKQIGVSGAGKSTILLSILGLINISIKVTEGDIRTYNRKNGDSIIGIQAQSFVYNNWIRFIDPYNFTDTEVHALKLGINLGNLDIKYHMKDQKSNYKIKITQSKVINGNSNDTILLNDICRIYSLVRLNLNRYKYLILDEIPILNPLNNSYBDELSFLIGKAKSFVITNRRHFNPTVLLISHAANTLSCDQYIVLRKGEITRYCSYEDVKTQSEHLSDMD"  
 rRNA 23896..31533  
 /gene="rRNA"  
 /note="region containing small subunit, 5.8S and large subunit rRNA genes and spacer regions"  
 gene 23896..31533  
 /gene="rRNA"  
 gene complement(31966..32775)  
 /gene="MAL1P3.04"  
 CDS complement(join(31966..32476,32675..32775))  
 /gene="MAL1P3.04"  
 /note="MAL1P3.04, conserved hypothetical membrane protein, len: 203 aa, similarity: p. falciparum chromosome 2, PFB0110W, O96126 predicted integral membrane protein (255 aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity in 191 aa overlap)"  
 /codon\_start=1  
 /product="conserved hypothetical membrane protein, MAL1P3.04"  
 /protein\_id="CAB63559.1"  
 /db\_xref="GI:6594247"  
 /translation="MKKSYPIWTVLILFLLLFLTYNNYDFSKTKFNNIKIDINRFRKILASAEQKQVPEEDNLIKNEELRIPENHNSPYLPERYENIDNELSINS TKIHETIKKRWQKTEYDNNHNVQDPWYKAWATYLYVHILINTILKDLTPNIDHETETITWIKWIGDIEYFLNGLVEMLALITLLEPYKNNK"  
 misc\_feature complement(32477..32486)  
 /gene="MAL1P3.04"  
 /note="potential splice acceptor sequence"  
 misc\_feature complement(32669..32674)  
 /gene="MAL1P3.04"

Query Match 3.48; Score 189.6; DB 60; Length 130281;  
Best Local Similarity 44.94; Pred. No. Be-14;  
Matches 979; Conservative 0; Mismatches 1179; Indels 22; Gaps 6;

Qy 1810 aaataattattataaaattatgatctgacttgctgactgtctgactaatcttcgacatttt 1869  
Db 98635 AAATAAATCTAATAAATAATTTTTTTTGTAGATATCTTCGATGAATATGAATATCT 98694

Qy 18070 aatttttttgggttggttgatatttgatatttgatatttgatatttgatatttgatatt 1929  
Db 98695 GTAAAAAAATATAATAAATAGATTTTAATTAATTATATTAATACTA-CAATAATTA 98752

|    |       |                                                                |       |
|----|-------|----------------------------------------------------------------|-------|
| Qy | 2321  | agggtgcagctacaacatttcccttttctccaaatacagatgttgatgcgaagtccg      | 2380  |
| Db | 17899 | ATTATATTATATATATTTAAATTTATATATTTGAATTTTATATATATATATA           | 17958 |
| Qy | 2381  | atctataaattattatcagtttttcaattccaaactacctatacctctattataaa       | 2440  |
| Db | 17959 | TATATGAGAAAATTAAATTTTAAATTTTAAATTAATTAATTTTTTAAAAATTTCTATA     | 18018 |
| Qy | 2441  | tataagtcagttccattcgtgttttcgaagttcccaaaatttgaattttataaattt      | 2500  |
| Db | 18019 | TGTATTATTTTATAAAAAATA TTATATATATAAATCATGTGTTTTAAAAATTAACAA     | 18078 |
| Qy | 2501  | attccctaaaaacgaattagtatactttcaaatttaagttctttttccattccgatt      | 2560  |
| Db | 18079 | AAAAATTTTAAATAAATAATTTTATAATGAATTAATTTTATTATTTTTCAAATTTTT      | 18138 |
| Qy | 2561  | tcaatttcactctttataactctctattctataataatacaaaattccaaataattt      | 2620  |
| Db | 18139 | AAAAAATTTTAAAAAATAATTTTTTTTTTAAAAAAAACATATACTCAATTATATA        | 18198 |
| Qy | 2621  | tgaattatctacattcgttccctaaagttccaaactataaatttcaacttgaagaattaa   | 2680  |
| Db | 18199 | TTATAGATATTATATATATAAATATTTAAATATTTATATATATCTATAAATTATA        | 18258 |
| Qy | 2681  | tcaattttccactcaagcatcaaaatttaaccaaatgacacaaatttcgtgattgtag     | 2740  |
| Db | 18259 | ATAAAAAATTTAAAAATTTAAAAATGTAGATATTTTATAAAAAATTATATCTCATAT      | 18318 |
| Qy | 2741  | atcagcttttgagctctcaaaacataaaattacaaaaaaacaaactaaatcat          | 2800  |
| Db | 18319 | TTATTATATTAAATTTAAATTTATATAAATATATAGATGATTAATTTATATATATAT      | 18378 |
| Qy | 2801  | ttatcatttgcagacaaacagctgtcccaagtcgaagcttcaaaatggctctcttg       | 2860  |
| Db | 18379 | TTATAAATTTATATATATATAGAATTTATATAATAATATATATATATATAGAAAAATTA    | 18438 |
| Qy | 2861  | ttctctttgtgcacacagctgagagagaagggaatgaagatgacacattttttta        | 2920  |
| Db | 18439 | ATTATTATAAATAATTTAATAAATTTTTAAAAATTTCTTAAGTGATTTTATTATA        | 18498 |
| Qy | 2921  | ttatgttttaacataataattataaattttt---aatcataattatacttggctgagtg    | 2976  |
| Db | 18499 | AAAAATTTTATATATAAATCATGTGTTTTTAAAAATAAACAAAAAATTTTATATAAT      | 18558 |
| Qy | 2977  | gacagctgggagctacgtcaagatttttaacctatactcttttgacagcgttggctgtg    | 3036  |
| Db | 18559 | AAATTTTAAATGAATATAATTTATTATTTTTTCAAATTTTTTAAAAAATTTTAAAA       | 18618 |
| Qy | 3037  | ctaccocagagtgctcaaaagtttgagctgcttcacagagcacttttggccataatgg     | 3096  |
| Db | 18619 | AAAAATAATTTTTTTTTTAAAAAAAACATATATCATTAATTAATAATTATAGATTTATA    | 18678 |
| Qy | 3097  | ataaagcgcaatttgttgatctcaactgctcacagaataatgtaaaatgaattaaacata   | 3156  |
| Db | 18679 | TATATATAAAATTTAATATATATATATATATATATCAATAAATTTAAATAAAAAATTTAAAA | 18738 |
| Qy | 3157  | agggtgctgtctcacacacacaaaaaaacattatgtgttggtggaattttatactac      | 3216  |
| Db | 18739 | TTTAAATGTAGATATAATTTTATAAAATTTATATCTCATATTTATTATTTAAATTT       | 18798 |
| Qy | 3217  | ggagctgaattattttttaaataaaattatgtttattagattcttaataatttttggag    | 3276  |
| Db | 18799 | TAAATTATATAAATAATATAATGATTTAATTAATATATATATTTTAAATTTATATAT      | 18858 |
| Qy | 3277  | cattccactactataatttctgatacaataataaaatatagataata-----           | 3324  |
| Db | 18859 | TATTGAAATTTATATATAATATATATATATATATAGAAAAATTTAAATTTTAAATATTT    | 18918 |
| Qy | 3325  | ----aagtgtaatttaacttttaaatcaacgacataataaaattttgaactcaattta     | 3379  |
| Db | 18919 | AAATATAAATTTTTAAAAAATTTCTTAAATGTATATTTTATAAAAAATTTTATATAT      | 18978 |
| Qy | 3380  | ttattctctcatatttttaacttaattatgcctattttttccaaataaaatttaactctaaa | 3439  |

|    |      |                                                              |      |
|----|------|--------------------------------------------------------------|------|
| Qy | 3440 | taaaataaactttttcttcaatgaatgcacactcgttatctactccaattatagatg    | 3499 |
| Db | 4123 | TATATATTTATTTATTTTCATTTTTTAAAAAAATTTTAAAAAAATATTTTTT         | 4182 |
| Qy | 3500 | tatatttaaccttgatgatttatcttatgtatattactctgtatataattgctggg     | 3559 |
| Db | 4183 | TAAGAAAAAATCATATCATATAAATTAATAGTATATTATATATATAAATATTTA       | 4242 |
| Qy | 3560 | tacactcgttttccataaatttttaactatgattataaattttttccacatcgat      | 3619 |
| Db | 4243 | ATATATATATATATATCTAATTAATTAATAAAAAATTTAAATTTAAAAATAGATA      | 4302 |
| Qy | 3620 | attactctattatacataattttctcatattttatgaaattgagacagaacaacata    | 3679 |
| Db | 4303 | TAATTTATAAAAAATTTATATCTCATATTTATTTATATTAATTAATTTATATAAATA    | 4362 |
| Qy | 3680 | agagacaacaattctatcacacagacaatttagaanaaaatgactctttagggtatttaa | 3739 |
| Db | 4363 | TATAGATTTTAAATTAATTAATTAATTTATTAATTTATATATTAATGAATTTATAT     | 4422 |
| Qy | 3740 | gtactcttaaccacacaacaaattccaattcaatgaactaactaagatatacatacata  | 3799 |
| Db | 4423 | ATATATATATATATATAGAAAAATTAATTTATTAATTAATTTTCATAAATTTAAAA     | 4482 |
| Qy | 3800 | cggaacactcttactctgtaactcttaactccatcaatttatatgaaaaatacttata   | 3859 |
| Db | 4483 | -AAATTTCTTAAGATATATTTTAAATAAAAAATTACTTTTTAAAAAAATATTTTAA     | 4540 |
| Qy | 3860 | tctctgaactaatggtgtgcacaaatttatctatcaataaagaaaaacacttaatttt   | 3919 |
| Db | 4541 | TTTTTAAAAAATTAAGTAAATATATAAAAAATAAAAAATAAAAAATGAAATTTATAT    | 4600 |

## AUTHORS

|            |                                                                                                                                                                             |          |     |          |     |             |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|-----|----------|-----|-------------|
| ACCDS      | DMW37541                                                                                                                                                                    | 19517 bp | DNA | circular | INV | 04-APR-2000 |
| DEFINITION | Drosophila melanogaster complete mitochondrial genome.                                                                                                                      |          |     |          |     |             |
| ACCESSION  | U37541                                                                                                                                                                      |          |     |          |     |             |
| VERSION    | U37541.1 GI:1166529                                                                                                                                                         |          |     |          |     |             |
| KEYWORDS   |                                                                                                                                                                             |          |     |          |     |             |
| SOURCE     | Drosophila melanogaster.                                                                                                                                                    |          |     |          |     |             |
| ORGANISM   | Mitochondrion Drosophila melanogaster                                                                                                                                       |          |     |          |     |             |
|            | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |          |     |          |     |             |
| REFERENCE  | 1 (bases 12511 to 12682)                                                                                                                                                    |          |     |          |     |             |
| AUTHORS    | Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.                                                                                                    |          |     |          |     |             |
| TITLE      | Drosophila mitochondrial DNA: a novel gene order                                                                                                                            |          |     |          |     |             |
| JOURNAL    | Nucleic Acids Res. 10 (21), 6619-6637 (1982)                                                                                                                                |          |     |          |     |             |
| MEDLINE    | 83090428                                                                                                                                                                    |          |     |          |     |             |
| REFERENCE  | 2 (bases 5269 to 5695)                                                                                                                                                      |          |     |          |     |             |
| AUTHORS    | Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.                                                                                                                          |          |     |          |     |             |
| TITLE      | Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes                                     |          |     |          |     |             |
| JOURNAL    | Nucleic Acids Res. 11 (8), 2411-2425 (1983)                                                                                                                                 |          |     |          |     |             |
| MEDLINE    | 83220794                                                                                                                                                                    |          |     |          |     |             |
| REFERENCE  | 3 (bases 404 to 5272)                                                                                                                                                       |          |     |          |     |             |
| AUTHORS    | de Bruijn,M.H.                                                                                                                                                              |          |     |          |     |             |
| TITLE      | Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code                                                                                            |          |     |          |     |             |
| JOURNAL    | Nature 304 (5923), 234-241 (1983)                                                                                                                                           |          |     |          |     |             |
| MEDLINE    | 83245048                                                                                                                                                                    |          |     |          |     |             |
| REFERENCE  | 4 (bases 804 to 1778)                                                                                                                                                       |          |     |          |     |             |
| AUTHORS    | Satta,Y., Ishiwa,H. and Chigusa,S.I.                                                                                                                                        |          |     |          |     |             |
| TITLE      | Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species                                                               |          |     |          |     |             |
| JOURNAL    | Mol. Biol. Evol. 4 (6), 638-650 (1987)                                                                                                                                      |          |     |          |     |             |
| MEDLINE    | 88174373                                                                                                                                                                    |          |     |          |     |             |
| REFERENCE  | 5 (bases 5268 to 13619)                                                                                                                                                     |          |     |          |     |             |
| AUTHORS    | Garesse,R.                                                                                                                                                                  |          |     |          |     |             |
| TITLE      | Drosophila melanogaster mitochondrial DNA: gene organization and                                                                                                            |          |     |          |     |             |



QY 3479 atacctcaaaatataagatataataattacctgatgatattatatttagtataatt 3538  
DB 138317 ATATAATAAATATTTATTAATTAAGTATATAATAATATTAATAATATAATAA 138258

QY 3539 ctgatattataatgatgggagatacatcgcttccncaaaatttaactatgattat 3598  
DB 138257 TATATATATATATATTTTAAATTTATATAAAATTTATATATTTTATATAATTTAA 138198

QY 3599 aaatttattccaaacgctatttactattataaacaaatttataacttaattataga 3658  
DB 138197 AAAATATATATATAATATATAAAAAATTAAGTATATAAAAAATATATATTTTATA 138139

QY 3659 aattgagaccagaaacatttaagagacaacaaattctatacaagaacatttagaaaaa 3718  
DB 138138 TTTTATAAATAAATAAATAAATATATAATTTAAATTTATATATTTTATATATTTTAA 138079

QY 3719 tgcatttttagtgatattttaaagcctcttaacaaacacaaatttcaactcaaatgac 3778  
DB 138078 ATTTATCATATAATATATTTTATATATAATTAATTTATATATATATATATTTTATATA 138019

QY 3779 taataagatatataataacacggagacacttctactgtactcttatactccataattt 3838  
DB 138018 AATTTATACATATATAATTAATTAATTTTATATATTTTATATAAATAGGTATTAAT 137959

QY 3839 attatgaaaaataacttatatacttcgaacaaattgtctcaacaaattatctcaaat 3898  
DB 137958 ATAAATAAATTAATATATATATATAGTATAAATATATATAATTAATTAATTTAA 137899

QY 3899 aaagaaacacacttaatttttatacaactttttcatataatttgaaagattatatttga 3958  
DB 137898 ATAAATATATATAAATTTTAAAAATATTTTAAATATATAAATATATATTTTAAAA 137839

QY 3959 tatttaagaaaaatttttgacatagatggacacctcttaa - caataccccacata 4016  
DB 137838 ATTTAATTTATATAAATTAATATGATATATAAATAAAAAAATTAATATATATAT 137779

QY 4017 agtccagtatgatagagaatttgtaacaacacgtggggccaaatcccccacacaca 4076  
DB 137778 TTAATATATATAAATATATATTAAGTATAAATAATTTTATTTATATGATTAATTA 137719

QY 4077 tctctctctctctctataaaagcgttgctcaactagacacaaatccacacaaatc 4136  
DB 137718 TAAATTAATTTATTTATATATGTTGTAAATTTTGTATTTATATATATACAAATA 137659

QY 4137 acgtctctttctcttatttgataaacagctgcataagatctgcacacctctctct 4196  
DB 137658 AATATGCAATATATTTTAACTGCGTATAAATAATAAATCACTCTCTATTTAAT 137599

QY 4197 ttccaactctttactcaaatgtctcaactagtcagccgtatgcacacgtgttcggcagc 4256  
DB 137598 CATTCAACGGTATATATAATATATATATCTTATATGCTCAAAAAATATATAGNTGA 137539

QY 4257 ggctcgacgttatctcgagacacaagaacacctctcaagactccccaaattggtccaa 4316  
DB 137538 TATTCAATATGATATATATATATGATATGTTTTNNNNNNNNNNNNNNNNNNNNNNNN 137479

QY 4317 atagaaagcagagagactctgaatacgaaaagccagatcaaacacagcaagaattcac 4376  
DB 137478 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 137419

QY 4377 aagagactcaaaaactctgagaagctgaaatgcaaaaggaggaacacaaaacactgca 4436  
DB 137418 TATAATATAAGTGATAGTTATATATTAACAATATATGATGTTGATATGATATATATA 137359

QY 4437 aacagactgaaggtaccacagctgcacgaactcaaaaggacaaaaggatcagcgaag 4496  
DB 137358 TATAAAATATATATATATATTAATATATATAAATGAATTAATTAATAGTATTTA 137299

QY 4497 aaactctgcagcaattccccccggctgcagcgctgacgagatcttcggcccgctcag 4556  
DB 137298 TATATATGATATAAAGATTAGTTATTAATTAATTAATATAAATAAATTAATATA 137239

QY 4557 ccttcagaatataacacact-gtcactctccacatcatcgcaatattctcagatctat 4616

Df 137238 ATATAAATATATATATAATCATCTGATTATATAAAATTCAGTATTATAACATGTAACTTAATTAAG 137129

Gy 4616 gcaataatgatatgttaataaaagattggtagtttgggaagtctgctgcacctcactccac 4675

Df 137178 TGTTAAAGACTATCAATTAATTAATATATATTAATATATAATATGTTTATTAATATGAATAT 137119

Gy 4676 gccaatgatggtgaaactcttggtcacatagaagaacttcasaagggttatagtttatgttat 4735

Df 137118 ATTATAT ---- TAAATGTATGTTATTGATGTTTTAGTATTTAATTAATATGAATAA 137063

Gy 4736 agtgttatggttagtgaanaatttaattaaagtctgtcatctaaggttaaocactaccgtgcc 4795

Df 137062 AAGTGATATATATTAAGAATATAAATATATATTTTTTATAAATAAATATAAAGAAA 137003

Gy 4796 tgattttctgtatgtatgtcttttacttaagtgetatgcagtattgcttaatttaaca 4855

Df 137002 GAATATATTTTTATATAGTATATATAGTATATATAGTGATTTTAAATATATAAATA 136943

Gy 4856 ttgctgtcatctatactctctcacataataataaaggcgactggtttgtttaaacct 4915

Df 136942 ATATAGCTTATATATATTTTTATATAGTATTATTAACAATATAGATATTATATAAAT 136893

Gy 4916 tttacnagttgaagcatgtataeatatagaacataataatcacggttttagtcagtgt 4975

Df 136892 ACTTAATATATAAATATATATATATAAATATAGTATAAATATATTTTTATAATTAGA 136823

Gy 4976 tagctctcttagtctgttatgtctgtctttaatcttaaccttaaacnaactccaactaaaat 5035

Df 136822 TAATTATATAAATTAATAAAAATATAGTATTATAGTATATATAT - AAGGAATATATAAT 136764

Gy 5036 tttatcaataataacnaacnaatattgttgtaataataacataactaaatgacaaaaaagta 5095

Df 136763 GHTTATATCTATTAATATATATATATATAAATATATATATATAAATTAATATATATA 136704

Gy 5096 aaataaaanaaaataagacaabaattgttatataattgtataataatgtacacatat 5155

Df 136703 ATATTAGATATATTTGGTATATATATAGATATATAAGACGAAATATCGATATATTT 136644

Gy 5156 tctcataactgaataggtt- ctacactataactccaactaaatttcagtttaaatatttt-- 5211

Df 136643 AGTATATATATGGGAGTAGATATANATATATAAGGTGATTTATAGTATATATAGTAG 136584

Gy 5212 -tatcccactaccatatattagaagactcttttttaasatataaaaattttatataccaat 5270

Df 136583 GAAATATATAGTTAAATATAGTATATATCTTAATATATATATATAGTATTATAGTGA 136524

Gy 5271 ttaatttaactaatatattctotctaactaaaactaaaattttattataoactatatataaa 5330

Df 136523 TTGAAATATATAATTAATATATATTAATTAATCAGAAAAATATATTGTGATAGTATTTT 136464

Gy 5331 ttccccattctcttatccatttaaacactctactatacccaatttaatttaaacctcttaa 5390

Df 136463 ATATATCATATAGTATTATATATATATAATAGTATNNNNNNNNNNNNNNNNNNNNNNN 136404

Gy 5391 ttatcttaatttgtaaoctoctocaaccagctagctgtcgagccagatccggagatta 5450

Df 136403 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATA 136344

Gy 5451 cat 5453

Df 136343 TAG 136341

```

RESULT 10
DMU1584
LOCUS      DMU1584      4601 bp      DNA              INV      23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION  U1584
VERSION    U1584.1  GI:508626
KEYWORDS   mitochondrial DNA; A+T region; tandem repeats.
SOURCE     fruit fly.
ORGANISM   Mitochondrion Drosophila melanogaster

```

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Mummatti, S.R., Varra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Locant, M.A., Wada, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.I. and Rubin, G.M.

## TITLE Direct Submission

JOURNAL Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

## COMMENT

On Mar 8, 2000 this sequence version replaced gi:7025688.

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently consists of 133 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 829: contig of 829 bp in length  
 \* 830 909: gap of unknown length  
 \* 910 1649: contig of 740 bp in length  
 \* 1650 1729: gap of unknown length  
 \* 1730 1953: contig of 224 bp in length  
 \* 1954 2033: gap of unknown length  
 \* 2034 2651: contig of 628 bp in length  
 \* 2652 2741: gap of unknown length  
 \* 2742 3349: contig of 608 bp in length  
 \* 3350 3429: gap of unknown length  
 \* 3430 3968: contig of 539 bp in length  
 \* 3969 4048: gap of unknown length  
 \* 4049 4504: contig of 456 bp in length  
 \* 4505 4584: gap of unknown length  
 \* 4585 5411: contig of 827 bp in length  
 \* 5412 5491: gap of unknown length  
 \* 5492 6373: contig of 882 bp in length  
 \* 6374 6453: gap of unknown length  
 \* 6454 7449: contig of 996 bp in length  
 \* 7450 7529: gap of unknown length  
 \* 7530 8139: contig of 610 bp in length  
 \* 8140 8219: gap of unknown length  
 \* 8220 9315: contig of 1096 bp in length  
 \* 9316 9395: gap of unknown length  
 \* 9396 10239: contig of 844 bp in length  
 \* 10240 10319: gap of unknown length  
 \* 10320 10964: contig of 645 bp in length  
 \* 10965 11044: gap of unknown length  
 \* 11045 11648: contig of 604 bp in length  
 \* 11649 11728: gap of unknown length  
 \* 11729 12696: contig of 968 bp in length  
 \* 12697 12776: gap of unknown length  
 \* 12777 13976: contig of 1200 bp in length  
 \* 13977 14056: gap of unknown length  
 \* 14057 15045: contig of 989 bp in length  
 \* 15046 15125: gap of unknown length  
 \* 15126 15969: contig of 844 bp in length  
 \* 15970 16049: gap of unknown length  
 \* 16050 16859: contig of 810 bp in length  
 \* 16860 16939: gap of unknown length  
 \* 16940 17662: contig of 723 bp in length  
 \* 17663 17742: gap of unknown length  
 \* 17743 18767: contig of 1025 bp in length  
 \* 18768 18847: gap of unknown length  
 \* 18848 19809: contig of 952 bp in length  
 \* 19810 19889: gap of unknown length  
 \* 19890 21046: contig of 1157 bp in length  
 \* 21047 21126: gap of unknown length

\* 21127 21826: contig of 700 bp in length  
 \* 21827 21906: gap of unknown length  
 \* 21907 23735: contig of 1829 bp in length  
 \* 23736 23815: gap of unknown length  
 \* 23816 25556: contig of 1741 bp in length  
 \* 25557 25636: gap of unknown length  
 \* 25637 26792: contig of 1156 bp in length  
 \* 26793 26872: gap of unknown length  
 \* 26873 28359: contig of 1487 bp in length  
 \* 28360 28439: gap of unknown length  
 \* 28440 29698: contig of 1459 bp in length  
 \* 29699 29978: gap of unknown length  
 \* 29979 31836: contig of 1858 bp in length  
 \* 31837 31915: gap of unknown length  
 \* 31917 33347: contig of 1431 bp in length  
 \* 33348 33427: gap of unknown length  
 \* 33428 34568: contig of 1141 bp in length  
 \* 34569 34648: gap of unknown length  
 \* 34649 35754: contig of 1106 bp in length  
 \* 35755 35834: gap of unknown length  
 \* 35835 37815: contig of 1981 bp in length  
 \* 37816 37895: gap of unknown length  
 \* 37896 39641: contig of 1746 bp in length  
 \* 39642 39721: gap of unknown length  
 \* 39722 41135: contig of 1414 bp in length  
 \* 41136 41215: gap of unknown length  
 \* 41216 42477: contig of 1262 bp in length  
 \* 42478 42557: gap of unknown length  
 \* 42558 44229: contig of 1672 bp in length  
 \* 44230 44309: gap of unknown length  
 \* 44310 45922: contig of 1613 bp in length  
 \* 45923 46002: gap of unknown length  
 \* 46003 47999: contig of 1997 bp in length  
 \* 48000 48079: gap of unknown length  
 \* 48080 49982: contig of 1903 bp in length  
 \* 49983 50062: gap of unknown length  
 \* 50063 51360: contig of 1298 bp in length  
 \* 51361 51440: gap of unknown length  
 \* 51441 53101: contig of 1661 bp in length  
 \* 53102 53181: gap of unknown length  
 \* 53182 54926: contig of 1745 bp in length  
 \* 54927 55006: gap of unknown length  
 \* 55007 56937: contig of 1931 bp in length  
 \* 56938 57017: gap of unknown length  
 \* 57018 57606: contig of 589 bp in length  
 \* 57607 57686: gap of unknown length  
 \* 57687 58632: contig of 946 bp in length  
 \* 58633 58712: gap of unknown length  
 \* 58713 60613: contig of 1901 bp in length  
 \* 60614 60693: gap of unknown length  
 \* 60694 62727: contig of 2034 bp in length  
 \* 62728 62807: gap of unknown length  
 \* 62808 65311: contig of 2504 bp in length  
 \* 65312 65391: gap of unknown length  
 \* 65392 66685: contig of 1294 bp in length  
 \* 66686 66765: gap of unknown length  
 \* 66766 68830: contig of 2065 bp in length  
 \* 68831 68910: gap of unknown length  
 \* 68911 71103: contig of 2193 bp in length  
 \* 71104 71183: gap of unknown length  
 \* 71184 72193: contig of 1010 bp in length  
 \* 72194 72273: gap of unknown length  
 \* 72274 74138: contig of 1865 bp in length  
 \* 74139 74218: gap of unknown length  
 \* 74219 76236: contig of 2018 bp in length  
 \* 76237 76316: gap of unknown length  
 \* 76317 77913: contig of 1597 bp in length  
 \* 77914 77993: gap of unknown length  
 \* 77994 80808: contig of 2815 bp in length  
 \* 80809 80888: gap of unknown length  
 \* 80889 82776: contig of 1888 bp in length  
 \* 82777 82856: gap of unknown length  
 \* 82857 85682: contig of 2826 bp in length



```

* be preserved.
* 1 67262: contig of 67262 bp in length
* 67263 67462: gap of unknown length
* 67463 82485: contig of 15023 bp in length
* 82486 82685: gap of unknown length
* 82686 130281: contig of 47596 bp in length.

FEATURES             Location/Qualifiers
     source            1..130281
                        /organism="Plasmodium falciparum"
                        /db_xref="taxon:5833"
                        /chromosome="12"
                        /clone="3D7"

BASE COUNT  52250 a 11780 c 11855 g 53996 t      400 others
ORIGIN

Query Match      3.6%; Score 197; DB 60; Length 130281;
Best Local Similarity 45.7%; Pred. No. 1.2e-14;
Matches 1118; Conservative 0; Mismatches 1290; Indels 37; Gaps 11;

Qy 1564 atgtgttttttgcgaattgttcacataatcatgcattcttcgaattgttcgaat 1623
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101320 ATTTTATTTATTTTATTAATAAAGGATAAAGACAATATAAATATATATATAAACA 101261

Qy 1624 tgaagcttaaaattctctatgatctgatctgttttaccattatattgtttgtatg 1683
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101260 TTCCAAATATATACCCCAATATATATATATATATATATATATATATATATATATATAA 101201

Qy 1684 ctgagtgtaagtcgaattcagattcagtcgaacccaattttattcatttcagga 1743
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101200 TATTAATATACACATATTAAT 101141

Qy 1744 atctgcgaactagggtgatgtggttcgagtcgagtcggtggttttccatccatcat 1803
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101140 AT 101081

Qy 1804 tttattaaataatttaattaaataatttgagctttggagctgctgcatttcag 1863
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101080 AAAATTAAT 101021

Qy 1864 aattttatttggttttggtttgttggaatttttagataatttttaattattctgc 1923
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101020 TAATTAATTTAAATATTAAT 100961

Qy 1924 ataattttctgttatttggaagagtcgttcgaatttttttccaaattgaacgttaa 1983
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100960 ATATTAT 100901

Qy 1984 gaatttttactactgcgaattcgaataagtggaattgttttttagaagaattaaag 2043
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100900 AAAT 100841

Qy 2044 ttgatttaccagttttt---agtttgatttggggaagtaatgtattgttttgacata 2100
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100840 TTAT 100781

Qy 2101 attatttgacaataatgaattttctagggaataacggaattctctctttttttgt 2160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100780 ATAAAT 100723

Qy 2161 aaattactaagtcgaagacacacacagcttttggggagcacaataatctagcttaagtag 2220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100722 TAATTAAT 100663

Qy 2221 tcaagttaactctcgaactctggtctacactctagcgt---tgagtttgctgtgacag 2277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100662 AATATTTAAATATTAAT 100603

Qy 2278 tagtagctctagaactctacgtcgaacacagcagctgagctgaggtcgactacacac 2337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100602 AAT 100543

Qy 2338 -ttttcttttttcttcagttacatactgttgatcgaagtcgactctatataattta 2395

```

```

      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100542 AAT 100483

Qy 2396 ttacgatttctacatttcactctctctctctctctctctctctctctctctctctct 2455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100482 ATATTTTAT 100423

Qy 2456 ttacgatttctcgaagtcgacacacacacacacacacacacacacacacacacacac 2515
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100422 TAAAT 100363

Qy 2516 aatagttatct 2575
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100362 AAT 100303

Qy 2576 tataact 2635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100302 AATTAT 100243

Qy 2636 ttgactctcgaagtcgaacacacacacacacacacacacacacacacacacacac 2695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100242 TAAAT---TAAAT 100185

Qy 2696 aagctcgaatttctcgaacacacacacacacacacacacacacacacacacacac 2752
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100184 TAAAT 100125

Qy 2753 ----agctctcgaac 2807
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100124 AT 100065

Qy 2808 ttgac 2867
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100064 TAT 100005

Qy 2868 ttgttcgaac 2927
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100004 TAAAT 99945

Qy 2928 ttac 2987
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99944 AT 99885

Qy 2988 gatacgttaagtttttaacattatctcttgcgaagtcggtcgtctaccacagag 3047
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99884 TTATTTTATTTAT 99825

Qy 3048 tgatcgaagtttgagctcctcctcgaagtcacacacacacacacacacacacacac 3107
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99824 AATGGAAT 99765

Qy 3108 ttgttgatttcacactcctcgaacacacacacacacacacacacacacacacacac 3167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99764 ATATGTTTAT 99709

Qy 3168 tcac 3227
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99708 AAT 99649

Qy 3228 ttattttt---taaaataaaatttgatttttagactcttaatttttgagcattccta 3285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99648 ATATATTTAAATTAATTAATTAATTAATTAATTTCTTCTTAATTTATTTTAAAT 99589

Qy 3286 ctataatttctgac 3345
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99588 ATTAAT 99529

Qy 3346 caagcgtcaattctaaatttttgactcattttttttctctctctctctctctctct 3405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99528 ATAAAT 99476

Qy 3406 agtctattttttcgaac 3465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT      6
AC005504
LOCUS       AC005504    104992 bp    DNA             HTG             01-APR-1998
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            *** 3 unordered pieces.
ACCESSION   AC005504
VERSION     AC005504.3  GI:4558564
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
ORGANISM    Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1 (bases 1 to 104992)
AUTHORS     Hyman,R.W., Pung,Z.L., Qin,F., Tanaki,T., Kurdil,O.B., Conway,A.B.
            and Davis,R.W.
TITLE       Plasmodium falciparum 3D7 chromosome 12
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 104992)
AUTHORS     Hyman,R.W., Qin,F., Pung,Z.L., Conway,A.B. and Davis,R.W.
TITLE       Direct Subgenomic Sequencing of Plasmodium falciparum Chromosomes
JOURNAL     Submitted (21-AGO-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337172.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces

```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 58642: contig of 58642 bp in length
* 58643 58842: gap of unknown length
* 58843 91011: contig of 32169 bp in length
* 91012 91211: gap of unknown length
* 91212 104992: contig of 13781 bp in length.
```

```

FEATURES
  source
    1. 104992
      /organism="Plasmodium falciparum"
      /db_xref="taxon:5833"
      /chromosome="12"
BASE COUNT  44286 a  9326 c  9564 g  41411 t  405 others
ORIGIN

```

Query Match 3.68; Score 197; DB 1; Length 104992;  
Best Local Similarity 45.78; Pred. No. 1.2e-14;  
Matches 1118; Conservative 0; Mismatches 1290; Indels 37; Gaps 11;

Qy 1564 attgtgtttttctggaasaaattgcctatacaatactgcctctccaaattttgttcaat 1623  
Db 72352 ATTTTATTTATTTTTCATATAAAAAGGATTAAGCAGTAAATTAATTTAATAATAAAAAACA 72411

Qy 1624 tgaacgtataaaattctctatgatatactgctgttaatacatatattgtgttatg 1689  
Db 72412 TTCCAAATATATACCCCCAAATATATATATATATGTATTAAGCTATACAAATATATAA 72471

Qy 1684 cttgagcttaagtcacaaacttgagattactagctaccaccaatttaattacttaccgga 1743  
Db 72472 TATATAATCACATATTAATAATAATATATTTTATTAATTTATAAAATTAATAAATAT 72531

Qy 1744 atctgcgaacttaggatggagggcttcaggagcttggaattgggttctccatcatat 1803  
Db 72532 ATATATATAATATATATATATATCAAAATTAATCATTTATAAAATTTATTTAAATATATT 72591

Qy 1804 tttattaaattatatttaattaaaattttggagctttggagctgtgctactaaattctgc 1863  
Db 72592 AAAATTAATATATATATATTTATTAATAATTAAGTAAATTTAATTAATAAATAAATA 72651

Qy 1864 aattttattttgggttttgggttttggatttttttagataattatttttaacttctctgc 1923  
Db 72652 TAATAAATTTAAATATTAATAATAATTAATTCATAATACACTTAATTAATAAATATGA 72711

Qy 1924 ataattttctgttatctggaagtgctgcgaatttttttccaaattgaaagcttttaa 1983  
Db 72712 ATATTATATAAATAATAATAATAGAAAATTTATTAACAAATTTAAATTTAATAAATAAA 72771

Qy 1984 gaatttttactactgcaaatctgagataagtggaatttttttttgaagataaataag 2043  
Db 72772 AAATTTATATTTTATAAATAAATAAATTAATAAATTTATTAATAATATATAATAAA 72831

Qy 2044 tttagattacagctttttt--agtttgattgggtggaagtaagttgtttttgacaata 2100  
Db 72832 TTATATATAATTAATTTAAATTAATTAATAAATAAATAACTAATATTAATAATAAA 72891

Qy 2101 attattttgacaaataaagttttctctgggaataacaggaatactctctctttttgt 2160  
Db 72892 ATAAATATAATAAATAAATTTTATTAATAATAAT--ATAAATATAATAATAAT 72949

Qy 2161 aaatttactactgcgaagacaaacacggttttggggagcaaaatactactgatttaagtg 2220  
Db 72950 TAATTAATAATATAATAATAAATTAATTAATAATAATAATAATAATAATTAATTAAC 73009

Qy 2221 tcaagtctaacctctcaaaactcgtggtcataactctcaggc--tgagttgtcgtgctacag 2277  
Db 73010 AATTTATTAATTAATTAATAATAATAATATATATAAACAATTAATTTATAAATTAAG 73069

Qy 2278 taqtaactgctatacaaaacttactgcacaaacacatcagctcaggtgcactctacac 2337

Y 3858 tatctatcgcaataaattgtgtccaaacttattatctataaataagaa - aaacacttaat 3915  
Db 1 TATTACTCGGCCAAATGTCTCCACAACATATATCCAAAAAAAAGGTGGAATATAT 60  
Y 3916 tttttaaacatttttttctattgtgaagattatattgtatattatgaagaaaaat 3975  
Db 61 TTTTATACATTTTTTTCATATTATTGCAGATTATATTTGTATATTACGTAAGAAAT 120  
Y 3976 ttgcacatagattgagacccttttttaacataatccacacaaatgaacagtatgagatgag 4035  
Db 121 TTGCATATGATTGACACCCTTTTACCATATGCCACCATANGTCAGATATGTAGATGAG 180  
Y 4036 aaatttgtaacaaacagctggggcgaataatcccaaaaacatctctctctctctata 4095  
Db 181 AAATGGTCAACAAACAGCTGGGGCGAAATCCACAAAACATCTCTCTCTCTCTCTATA 240  
Y 4096 aaagggttgtaacacatgacacaaatccacacacaaatacac-----gttttttttt 4150  
Db 241 AAAGGCTAGTTCACATACACACATCCACACAAATTCACCTCAAAATTTTGTGCTT 300  
Y 4151 ctattt-gattaacatgtgctatgactgttcaccccttttctcttttccacatttta 4209  
Db 301 GTATTGTGGTTACCATGGCTCATACACTGTGCACCTTTTCTCTCTTCCACATTTTA 360  
Y 4210 ctctaatggtctctactagtagcaggtagacacactgcttoggcagggctgtagcttta 4269  
Db 361 CTCATTAGTGTCTCTCAATGAATCGGTAGCACACCGCTTCGCAGACGGCTCGAGCTTA 420  
Y 4270 ttgcagacacagcaacctctatcagaggtcccaacttgcttcaaaatcagaagacg 4329  
Db 421 TTGCAGACACAAACACCTCTACGTAGTGGCCACATATAGCTTCAAAATCGAAGAACAG 480  
Y 4330 -agagctgaattacgaagaagccgaatacaaaacagccaaggtctcagaagagttacta 4387  
Db 481 GAAGAGCTCGAATTGTAAGAGCGCGAATCAACACGCCAAAGTATGACGAAGATACCCA 540  
Y 4388 aaacttgagaagcctgaatgtcaaaaggaggaaaaaacaacacactgcgaacagatgaa 4447  
Db 541 AAACATGAGAAAGCTTGAATTCACAGAGAGGAAAAACAAAACCTGTCAGCACCATGAG 600  
Y 4448 gagtacacagagttacacagatcaaaaggagcaaaagatctcagagaagaaattctgaa 4506  
Db 601 GAGTACACAGGTCTACACAAATCGAAGAGCAGCAGAGATCTCCGAGAAGAAAAACCGGA 659

|           |                                                                                                                                                                                                      |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACCESSION | U34401                                                                                                                                                                                               |
| VERSION   | U34401.1 GI:1143223                                                                                                                                                                                  |
| KEYWORDS  | .                                                                                                                                                                                                    |
| SOURCE    | sea-island cotton.                                                                                                                                                                                   |
| ORGANISM  | Gossypium barbadense                                                                                                                                                                                 |
|           | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Malvales; Malvaceae; Gossypium. |
| REFERENCE | 1 (bases 1 to 1699)                                                                                                                                                                                  |
| AUTHORS   | Rinehart,J., Petersen,M. and John,M.E.                                                                                                                                                               |
| TITLE     | Tissue-specific and Developmental Regulation of Cotton mRNA, FbLate-2: Promoter Studies in Transgenic Plants                                                                                         |
| JOURNAL   | Unpublished                                                                                                                                                                                          |
| REFERENCE | 2 (bases 1 to 1699)                                                                                                                                                                                  |
| AUTHORS   | John,M.E.                                                                                                                                                                                            |
| TITLE     | Direct Submission                                                                                                                                                                                    |
| JOURNAL   | Submitted (21-NOV-1995) Malayakal E. John, Fiber Technology, Agracetus, 8520 University Green, Middleton, WI 53562, USA                                                                              |
| FEATURES  | Location/Qualifiers                                                                                                                                                                                  |

Query Match 4.8%; Score 265.4; DB 5; Length 1283;



Query Match 2.1%; Score 115.2; DB 120; Length 1187;  
Best Local Similarity 47.0%; Pred. No. 2.9e-08;  
Matches 362; Conservative 0; Mismatches 402; Indels 6; Gaps 2;

CNS00210

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

|            |                                     |     |      |       |            |  |
|------------|-------------------------------------|-----|------|-------|------------|--|
| FEATURES   | Location/Qualifiers                 |     |      |       |            |  |
| SOURCE     | 1..1101                             |     |      |       |            |  |
|            | /organism='Drosophila melanogaster' |     |      |       |            |  |
|            | /db_xref='taxon:7227'               |     |      |       |            |  |
|            | /clone_lib='RPCI-98'                |     |      |       |            |  |
|            | /clone='BACR05W11'                  |     |      |       |            |  |
|            | /note='end : TET3'                  |     |      |       |            |  |
| BASE COUNT | 631 a                               | 7 c | 28 g | 289 t | 146 others |  |



... ..

Db 807 AAATATTTAATTTTTTAAATATATTTTATTTTNTTATTTTTNTTTATATTTTTTTATATT 866

```

RESULT 4
CNS0167M/c
LOCUS      CNS0167M      1201 bp      DNA      GSS      26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genome survey sequence.
ACCESSION  AL106396
VERSION     AL106396.1 GI:5621701

```

## ALIGNMENTS

Query Match 2.6%; Score 145.6; DB 122; Length 1101;  
Best Local Similarity 38.7%; Pred. No. 9.2; b3  
Matches 249; Conservative 130; Mismatches 260; Indels 5; Gaps 2;

Qy 3326 taacttaataatcaagacataataataattttgtagcatataatttttttctatatt 3385  
Db 457 TCAATTTGGGGGGGAAATATWAAAGAAATWTAATWAAAGAAAGWAAATTTTGGW 516

Qy 3386 tttataataatttagctctatctttttcaaaataaaatttaaatctaaataaaattttt 3445  
Db 517 WTATTTTGGGGGAAATATWAAAGAAATWTAATWAAAGAAAGWAAATTTTGGW 576

Qy 3446 octtaattgttgacaacactcgtgtactctcaaaattataagattt-attattctgt 3504  
Db 577 AAWATATTATATWATAAATWTWATTTATATAAAAAAATTTTWTWAAAAATTTTA 636

Qy 3505 atgattttttttattagatatattattcgtattataattcgtgggatacaactcgttc 3564  
Db 637 ATATATTATTATWTAATTAATAATWTAATWTWATTTATATAAATTTWAAATAAAAA 696

Qy 3565 cactaataattttaactatgattttataattttattcaacatcgtatttttactattaa 3624  
Db 697 AAAAAAATAATWAAATWATAATWATAATWATAATWATAATWATAATWATAATWATA 756

Qy 3625 tacataattttatcataatttttggaaattgagacagaacaaacttaagagacaattc 3684  
Db 757 TWTATATTATWTTTAAWATWAAATWTAATWATAATWATAATWATAATWATAATWATA 812

Qy 3685 tatacaagaacattagaaaaaaattctactcttttagtgatttttaagtactcttaacca 3744  
Db 813 ATATWATAATWATAATWATAATWATAATWATAATWATAATWATAATWATAATWATA 872





```

RESULT 10
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Peeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

```

Query Match 1.5%; Score 85; DB 1; Length 5852;  
Best Local Similarity 46.18; Pred. No. 1.2e-06;  
Matches 403; Conservative 0; Mismatches 460; Indels 12; Gaps 3;

Qy 3112 acgtgctccagcgaatagtgtaaaatgaatataaaatagtgccgtgctccacacaca 3171

Db 1381 AGTCAAGCACTAAATATGTATAAACCAAAAAATGTTAAGATGAAGTTATATGTTGT 1440

Qy 3172 aaaaaaactaatgtgtgtggtgtgatttttatatagcgaagtgaatatattattaaa 3231

Db 1441 ATTATATAAATAGTGTAGTTAAATTTTTATATCATTTTTTAAAAATGAAATGTTCG 1500

Qy 3232 ataaattatgtttattagatcttaataattttggagatctccatctataatttcgtac 3291

Db 1501 AAAAAAAAATTTTTTTTTTTTTTTCACGGCAGATGTAATATCATATGATCAAAAT 1560

Qy 3292 ctaataattaaatatagttataaagtgttaatttaactttaaattcaagcagataatt 3351

Db 1561 TAAAGATTATACCAATATGTAAATAATATAAAACATAACTAGTTATAAATCTTTC 1620

Qy 3352 aaatttggatcaattattttt---attctatatttttaatttaatttagctattttt 3408

Db 1621 CCTCTTTTTTTTTTTTTTTTGTGACAGCACTTTTTTTTTTGTGACAGCACTTTT 1680

Qy 3409 tcaaaataaaattcaatcaataaaataatttttccttaagtgtcaaacacatcgtg 3468

Db 1681 TTTTAAAAAAAATAAAATAGTTAAATACTATTGTAGACATCATTTTTCCTAGTT 1740

Qy 3469 tataattcaaaataaagtattatattaccctgatgtattattattagat---attat 3526

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826P  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH21.001CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
OS-08-487-826P-13

Query Match 1.88; Score 100.4; DB 4; Length 19124;  
Best Local Similarity 44.18; Pred. No. 3.1e-09;  
Matches 877; Conservative 0; Mismatches 1091; Indels 22; Gaps 10;

Qy 1962 attgaacgtttaagaatttttactactgcaaatccgaataagtgattgttttag 2021  
Db 4660 ATTCATATTTAGAGATTATGTAATATGTTTATATGCTAATATATATTAATTA 4719  
Qy 2022 aaagattaataatgttagttatctacgttttttagtttgattgggtgaagtaagtgt 2081  
Db 4720 GTTTTTTATGATGATAGTATGTTCTAATAATATATCATATAGTAGTCATAGTGTCA 4779  
Qy 2082 ttttgaacataatttttgacaataaattgaatttttaggtgaataacggaataattctc 2141  
Db 4780 ATATAAATATGCTATATTTATATATGTTATATATTAATTAAGTAAACACAGA-ACA 4838  
Qy 2142 ttcttttttgaataactaactgaacgaacaaacacgttttggggagcaataactcta 2201  
Db 4839 TATATAGTATAAATAGAGAAATATATATTTTATGCTATATATATAGTATATATA 4898  
Qy 2202 gotttaagtagtcagtgtaactctcaaaactcgtgataactcttgctgagtggtgtg 2261  
Db 4899 AAGGGCAAAATCATATATTTATGAAATTTTGTATATGATATAGTTAAGTAAAA 4958  
Qy 2262 tgctacagttagtgctatgaacttacctgacaaacagacagtgcgtcaggtgcag 2321  
Db 4959 AAAAAAAGAACAGCAAGCAAAATGGAAGCATAAAAATTTTACTGTATAGGATAAAA 5018  
Qy 2322 tctacaaatttctcttttcttctcaattacatagtgttgatcgaatttcgactataat 2381  
Db 5019 TATATATATAAATGTTTATTTATCTTAAAAAGTTCCTATATACATTAAGAAAA 5078  
Qy 2382 aattattacagatttatcaatttcaattcttatatcatctcttataaataagtc 2441  
Db 5079 TTGTCCTCATTTTATAAATTAATTAACACTATTACATATGAATTCGATTTGTGTTT 5138  
Qy 2442 gtccaaatcgtttttgcgaagttcccaaaattttgaaatttataaatttatccctaa 2501  
Db 5139 TTTTGATGAATATATGGAATATATTTATATGATGAGTGGTCTATATATAATATA 5198  
Qy 2502 aacggaatagttatctcttcaatttaagtttcaatttcaactcagttt-caatttc 2560  
Db 5199 ATTTTATTTAAAAAATGAAAAATGAAGAATAAATATCTCGATTTTGTATGCCATAGC 5258  
Qy 2561 atctttttatattctctattattctataactataactaaattcaaaatttaatttgaa 2620  
Db 5259 TTAATATATATGAGCTCATATATATATATATATCTTTACACAGCATATAGTAA 5318

Qy 2621 ttacacttttagtccctaggtccaacactaaattttcactttagaataatacatttt 2680  
Db 5319 ATATATTTTAACTCTATAAGGAAATAAAAATATAAATAAGAA---TACTGATAAA 5375  
Qy 2681 cacatctaaagctcaaatttacccaatgacaaatttctgattagttagtcagct 2740  
Db 5376 TAGTCATATATACAAATTTTAAAAATGTACATATATACAAATAGTACATGATATA 5435  
Qy 2741 tttagctcttccaacataaaaattcacaaaaaaacaaacttaaatctttatcaat 2800  
Db 5436 TAGAATATATAGAATTTATATTAAGGATAAATATAATATTTAAAAATATATTTTT 5495  
Qy 2801 ttgaacacaaagcttgccgaatgctaaagctttaaagcttctttgtttcttt 2860  
Db 5496 ATGTCAATTTATGTTATATATATATATTAACATGATTA-GTTTTTGA AAAATATTA 5554  
Qy 2861 tgttgcaacggtgagagagagaggggaaggaagattgacatattttttattattgt 2920  
Db 5555 AATATCATATAATAATAAATAGTTAAATAATAGTATTTATCAAAAATACTACT 5614  
Qy 2921 taacataataataataaatttaactaataattcttggtgaatgagacagtgaggag 2980  
Db 5615 TATAGTATATCATATAATAATATATATATATATTTATGCTGTTTTGATGGGTAT 5674  
Qy 2981 atactgaagatttttatactatacttttgcgaagcttgctggttaccocagag 3040  
Db 5675 ATAGGCTATAAGTATATATGGGTGTCATATATTTATATGGAATAGATACATAT 5734  
Qy 3041 gatcaca-----gtttgactgcctcaatgagccaattttggccataatgataaa 3093  
Db 5735 AAGTAAATATATTTTGTGATATGTCGTGTTAAGATAGATGCTATACAGTAAAG 5794  
Qy 3094 ggcgaatttgtttagtgcactgctcagacgaataagtttaaaatgaataaataaggt 3153  
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTTGTACATATATAAAAAATAGATACTAACATA 5854  
Qy 3154 gctgtgtcacacacacacacacacacacacacacacacacacacacacacacac 3213  
Db 5855 TCGATATACAGAATAATATTTGTATAAATATATATATATATATATAAAGACAT 5914  
Qy 3214 gtaatatataatttaaaataaattgtttatttagattcttaatttttggagcattc 3273  
Db 5915 -TAAACATACATAAGTAGTAATAGTTTATATATCATCTCTTTATATATAATTT 5973  
Qy 3274 catactataattctgtaacataataataataatagtaataaagtgtaataacttta 3333  
Db 5974 TTTTGTTTACTCTGTCGTCCTTTTTTTGTATATAATAAACAATATAAACAATA 6033  
Qy 3334 aattcagaactaatataaatttgaactaattattttattctatttttaata 3393  
Db 6034 TCGATATTTGGAATATAAATAAATTTATCTACATATAGCATATATATATATA 6093  
Qy 3394 atttagctatttttccaaaataaaatttaactcaaaaataattttt---cotta 3450  
Db 6094 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 6153  
Qy 3451 atgtgaaacaaactctgttactctcaaaattataagattatatttaccgtgatt 3510  
Db 6154 ATTTTATATATTTTAGTATATCTTTAAGATATTTATTAATATATATATAGCATAT 6213  
Qy 3511 tatttatagatataatacttgattataattgattgggatacaactgcttccactaa 3570  
Db 6214 ATGTATTTATATATACAAATATTTTCAATTTATATAATAATAGAACTAGAACATTT 6273  
Qy 3571 atattttaactatgatt-tataaatttatttcaactgattattactatttaatacat 3629  
Db 6274 ATTAATACACTCATATTGAATATATATTTATATAGTGTATTTTACTATTTTTTAT 6333  
Qy 3630 aattattcaaaattttatgggaattgagacccaagaacataagagaacaaattctata 3689  
Db 6334 ATTAACAATATAATTTGAATTCATAAAATGATGAATACATAAAAAATACACAA 6393

Qy 4244 cgagcgcgcgcagcttatctgagacaacagcaacctctacagagctccacacatttgct 4303  
Db 133 TCAGCGCTCGCATTATTATCAGACCAACACCTCTCTAGAGCTCGCCHATTGCT 192

Qy 4304 tcaaaatagcaaaagcgcagagagctctgaatcgaagaaagcgaatacaaacagcgaag 4363  
Db 193 TCAAAATAGCAAAAGCACAAGAGTCT-----GAATCAACACACAAA 237

Qy 4364 tatcgcagagagactcctaaacttgagagcctgaattgcaaaaggagaanaacaaa 4423  
Db 238 TATCAGCAAAATACCAAAATCATGAGAGCCTAAATGCACAGGAGAAAAACAAA 297

Qy 4424 cccctgcaaacgcgtgagagtagtaccagcagtcacagcgaatcagaagcgaagagatc 4483  
Db 298 CCCCTGCAACATCATGAGAGTACACAGCTACACGCGAATCGAGGAGCAGCAGAGTAC 357

Qy 4484 gagaagaanaattctgcagggccgcga 4510  
Db 358 GATAAGAAAAACCGAGTTTCCGCGAA 384

RESULT 5  
 US-08/787-335-18  
 Sequence 18, Application US/08787335  
 Patent No. 5961834  
 GENERAL INFORMATION:  
 APPLICANT: John, Maliyakal E.  
 APPLICANT: Umbeck, Paul F.  
 APPLICANT: Brill, Winston J.  
 TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS  
 TITLE OF INVENTION: FOR ALTERED FIBER  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Quarles and Brady  
 STREET: P.O BOX 2113  
 STREET: FIRST WISCONSIN PLAZA  
 CITY: MADISON  
 STATE: WISCONSIN  
 COUNTRY: U.S.A.  
 ZIP: 53701  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette - 3.50 inch, 800kb storage  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Macintosh  
 SOFTWARE: Microsoft Word 4.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/787,335  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/530,797  
 FILING DATE:  
 APPLICATION NUMBER: US 07/253,243  
 FILING DATE: 04-OCT-88  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Nicholas J. Seay  
 REGISTRATION NUMBER: 27,386  
 REFERENCE/DOCKET NUMBER: 1122990245  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1283 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: no  
 ANTI-SENSE: no  
 ORIGINAL SOURCE:  
 ORGANISM: Gossypium hirsutum  
 STRAIN: Coker 312  
 DEVELOPMENTAL STAGE: 15 day old fiber cells  
 TISSUE TYPE: fiber cells  
 IMMEDIATE SOURCE:

; LIBRARY: CKFB15  
; CLONE: E9  
US-08-787-335-18

```

Query Match      5.04; Score 273.4; DB 4; Length 1289;
Best Local Similarity 8.24%; Pred. No. 6.1e-39;
Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

Qy 4124 aatacacgctgtctttctttcttatattgataaacatggcgctacattgctacacctttc 4193
Db      13 ACTAATAAAATCTTCTGCTTCTATTTTGTAACCATCGGCTCATACTTTGTCAATCTCT 72

Qy 4184 ttctctttccacctttactactaagtgtctactagtcacgggtaggccacactgtttg 4243
Db      73 TTCTCTTTTCACCTTTTACTACTTACGTCTCACTAATAATCGGGTAGTCACACGCTCTG 132

Qy 4244 gcagcggctgcagctttattgtgagacacgaagcaactctatcagagctcccaactggtc 4303
Db      133 TCAGCGGCTCGACATTTATTCAGACACAAACAACCTCATCAGAGCTGCCAATTTGGCT 192

Qy 4304 tcaaaatgcgaagacgcgaagagtctgaattgcgaagaagcgaagaatacaaacgcgaag 4363
Db      193 TCAAAATACGAAGAAGCACAAGATCT-----GAATCAAAACACACAAA 237

Qy 4364 tatcgcgaagagtactccaaaattgaaagagctggaatgcgaagaggagaaacaaana 4423
Db      238 TATCAGCAAAAGTACCCAAAATCATGAAGAGCTGTAATAGTCACAGGAGGAAAACAAAA 297

Qy 4424 cctgcacacagcatgaagatatacagcaggtcacgcgaatcaaaaggcgaaaagatgac 4483
Db      298 CCTGCACACATCATGAAGAGTACACAGAGTCACGCGAATCGAAGGAGCAGGAAGATC 357

Qy 4484 gagaagaanaattctcgacgggcgcgaa 4510
Db      358 GATAAGAAAAACCCGATTCTCCCAAA 384

```

```

RESULT 6
US-08/487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Kin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NH121.001CP1
; TELECOMMUNICATION INFORMATION:

```





Db 5319 ATATATTTTAMCTTAAATAGAGAAATAAAATATTAATATAGAA--TACTGATAA 5375  
Qy 2681 cacatctgaagcatcaaatttacccaatgacacaaatttcattgattagctgaagct 2740  
Db 5376 TAGGCAATATATACATTTTAAAAATGTAACATATATACAAATAGTACATATATTA 5435  
Qy 2741 ttgagctttccaaacataaaaaattccaaaaaaacaaattcaaatctttatcaat 2800  
Db 5436 TAGAATATATAGAAATTTAATATAGGATAAATATTAATATTTAAAAATATATTTTT 5495  
Qy 2801 ttgaacacaaagcttgccgaatgctgaagccttaaaatggctcttttttctcttt 2860  
Db 5496 ATGTCAATTTATGTATATATATATATATATACATGATTA-GTTTTGAAAAATATTA 5554  
Qy 2861 tgttgcaaacgctggagagaagggaaatgaagattgacctattttttattatgttt 2920  
Db 5555 AATATCATATATATATAAATAGTAAAAATAAGTATTTTATACAAATACTAACT 5614  
Qy 2921 taactataatattcaataatttaacataataattcttggctgaatgacagctggag 2980  
Db 5615 TATAGTATATCATATAAATATATATATATATATATATATATATATATATATATAT 5674  
Qy 2981 atacgtaaagatttttaacataatactttttgcaagcagctggctgtccaccaagagt 3040  
Db 5675 ATAGGCTATAGATATATATGGTGTCTTATATATATATATATATATATATATATAT 5734  
Qy 3041 gatcaaa-----gtttgagctgcttcaatgagccaattttgcacataatgataaa 3093  
Db 5735 AAGTAAATATATTTATTTGTATATGCTGTGTTAAGATAGATATGATACATTAAG 5794  
Qy 3094 ggcatttggttttagttcaactgctcacagaataatggttaaatgaattaaataaggtg 3153  
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTTGTACATATATATATATATATATATATATAT 5854  
Qy 3154 gctgtgcacacacacacacacacacacacacacacacacacacacacacacacac 3213  
Db 5855 TGCATATACAAAGATAAATATTTGTATAAATATATATATATATATATATATATATAT 5914  
Qy 3214 gtaatatatttttaaaataaattatgtatttagatttcttaatttttggagcttc 3273  
Db 5915 -TAAATCTACTACTAGGTAAATAGTTTTATATATATATATATATATATATATATAT 5973  
Qy 3274 catactataatttgcacacataataatttaaaatagtaataaagtgtaattctta 3333  
Db 5974 TTTTGTATTTCTCTGTCTCTTTTTTGTATATATATATATATATATATATATATAT 6033  
Qy 3334 aattcacagcataattataaatttgaatcaattattttattctattatttttaatta 3393  
Db 6034 TCAGTATTTGGAAATATAAATAAATTTATCTCATATATATGATATATATATATATAT 6093  
Qy 3394 atttagctctatttttcaaaaataaaatttaaaatcaaaaataaaattttt---ccta 3450  
Db 6094 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6153  
Qy 3451 atgtgaaacactcagtttatacttcaaaatataagattattatttcaactgtgatt 3510  
Db 6154 ATTTTATATATTTTATATATATATATATATATATATATATATATATATATATATAT 6213  
Qy 3511 tatttattagatatataattctctgattataaattatggtgggatacaatogcttccactaa 3570  
Db 6214 ATGTATTTATATATACAAATATTTTCATTTATATATATATATAGAACATGAACATTT 6273  
Qy 3571 atattttaactatgattt-tataaattttattccacatogtataatttactattataaact 3629  
Db 6274 ATTAATACCTCATATTTGAATATATATATATATATATATATATATATATATATAT 6333  
Qy 3630 aatttatcatatttttgaagaattgagacgaacacataagagacacaaactatat 3689  
Db 6334 ATTATACATAAATTTTGAATTCATAAATGCATGAATATATATATATATATATATAT 6393  
Qy 3690 caagacaaatttgaacaaat-gtaacttttagttaatttttagtaactcttaaccacaa 3748

Db 6394 AACAAATGATAAAACATTTTTTATTAATATAATATAATATAATATAATATATTTTC 6453  
Qy 3749 caaaatctcaatcaaatgacataaagataataatacaatcaggaactcttactgt 3808  
Db 6454 CTGTATTTATTTATCATTTTTTTTTGATGCTATATATATATATATATAAATATTA 6513  
Qy 3809 aacttcaactcccaataattttatatagaanaatactttatataactcagactaaagt 3868  
Db 6514 ATATATA---ACACAAAATTAATATATATATATATATATATATATATATATATAT 6570  
Qy 3869 tgcacaaattatttctcaataaagaacacactaaatttttaacatttttctatat 3928  
Db 6571 TACAAGAAATATGATATATATATATATATATATATATATATATATATATATATAT 6630  
Qy 3929 atttgaaga 3938  
Db 6631 TATAGATAGA 6640

Search completed: September 3, 2000, 03:09:58  
Job time: 28544 sec

Qy 3854 actgaaacaaagttgtgcacaaattattatctaaaataagaaaaacattaattttat 3913  
Db 5559 ATATTTAATATTTTCAAAAACACTATCATGTTAATATAATAATA-TAACATAAATT 5501

Qy 3914 aacattttttctatattttgaagatattattttgatatatttcgtaaaaatttt 3968  
Db 5500 GACATAAAAAATATAATTTTAATATTTTATATTTATCTGTTAATATAAATCTT 5446

## Qy 343

```

Query Match          1.81; Score 101; DB 1; Length 3701;
Best Local Similarity 47.4%; Pred. No. 6.1e-06;
Matches 398; Conservative 0; Mismatches 435; Indels 6; Gaps 3;

Qy 3200 ttaataacggaggaatgaataatttttaaaataaaatgattgattagaattcaat 3259
Db 369 TTAATATCACTGGTATAATATTAAGAATATATAGTTTGTGATATGTATGATATA 428

Qy 3260 attttggagcatccatacataaatt----tcgtacataataattaaaataatgataat 3315
Db 429 AATTAGACAGATATAATATTAATATCAATATTATATTAATTTATTATATAGATTAT 488

Qy 3316 aaagtgttaata-actttaaatcacagcataaataaatttttgatcaataatttttt 3374
Db 489 TATTGATATTATATAATATCATATGTTGATGTGCATTATAGTATATATATCAATAT 548

Qy 3375 atttcataattatttaataatttaagtcatatttttccaaataaaatttaactaataaa 3434
Db 549 CCAATATATATATAATAATATGAAATTATTAATTAATGTCATATACATTATATAAT 608

Qy 3435 aaataattttctcctaagtgttgaaacacactgttatctactccaataataagattat 3494

```

RESULT 10  
 T70055  
 ID T70055 standard; cDNA; 1283 BP.  
 AC T70055;  
 DT 20-AUG-1997 (first entry)  
 DE Cotton fibre specific cDNA clone E9.  
 KW cotton; E6; fibre; promoter; transgenic plant; truncated;  
 KW heterologous gene expression; ds.  
 OS Gossypium hirsutum strain Coker 312.  
 PN US5620882-A.  
 PD 15-APR-1997.  
 PF 04-OCT-1988; 253243.  
 PR 04-OCT-1988; US-253243.  
 PR 21-NOV-1990; US-617239.  
 PR 18-MAY-1992; US-985970.  
 PR 19-OCT-1994; US-298829.  
 PA (CFTU) AGRACETUS INC.  
 PI John M;  
 DR WPI: 97-235185/21.  
 DT DNA constructs constg. truncated promoter sequence - for  
 PT fibre-specific gene expression in cotton plants  
 PS Example 3; Column 45-48; 48pp; English.  
 CC T70040-57 are cotton fibre-specific cDNA clones which can be used to  
 CC obtain genomic clones containing fibre-specific promoters. Claimed DNA  
 CC constructs comprise a truncated promoter sequence (from one of T70031-38)  
 CC that promotes preferential gene expression in plant fibre cells, a  
 CC protein coding sequence not naturally associated with the promoter  
 CC sequence and a 3' termination sequence. The DNA constructs are useful for  
 CC expressing foreign genes in fibre-producing plants, esp. to produce  
 CC transgenic cotton plants with varied cotton fibre characteristics and  
 CC quality. The present sequence comprises E9 cDNA isolated from clone  
 CC CKP815-E9 (CK = Coker; FB15 = 15 day old bolls).  
 SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 5.04; Score 273.4; DB 1; Length 1283;  
 Best Local Similarity 84.24; Pred. No. 2.4e-28;  
 Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

Qy 4124 aatacagctttctttctttctattgattacacgtgctcatgactctgctcattc 4183  
 Db 13 ACTAATAATTCCTTGGCTTCTTATTTGTAAACCATGGCTCATAACTTTTGTCACTCTTC 72  
 Qy 4184 tcttttttccaaactttactcaatgaagttctactagtagcagctgagccactgtttg 4243  
 Db 73 TCTCTTTTCCAACTTTTACTCATCTGCTCTCAATAATCGGTGATCAACCGCTCTG 132  
 Qy 4244 gacgagctgtagctttatttagacacagcaacccctcatgagctcccaactgct 4303  
 Db 133 TCAGCGGCTGAGCATTTATTCAGACACAAACCTCATAGAGCTGCCAACAATGGCT 192  
 Qy 4304 tcaaaactcgaaagacgacgagagctgtaatacgaaagcgaatacaaaaagccaaag 4363  
 Db 193 TCAAAATACGAAGACCAAGAGCTCT-----GAATACAAACCAACAAA 237  
 Qy 4364 tatcacgagagtagtcaaaacttgagagcctggaatcgaaagaggaatacaaaa 4423  
 Db 238 TATCCGAAGAGTACCAAAACATGAGAGCCCTAAATGTCACAGGAGGAAAAACAAA 297  
 Qy 4424 cctgcagacagactgaagtagtaccacagtagtaccagtagtcaaaagagcaaaagtag 4483  
 Db 298 CCTGCAAAATCATCATGAAGATACACGAGCTACCGGANTGAGGAGCAGCAAGAGTAC 357  
 Qy 4484 gagaagaatactctgacggccgca 4510  
 Db 358 GATAAGAAAAACCGGATTTCCCAAA 394

RESULT 11  
 T43361  
 ID T43361 standard; cDNA; 974 BP.  
 AC T43361;

DT 11-MAR-1997 (first entry)  
 DE Cotton FbLate 2-82A gene cDNA clone A8 (FbLate-1).  
 KW FbLate; promoter; fibre; transgenic plant; cotton; ds.  
 OS Gossypium hirsutum.  
 PN W09639021-A1.  
 PD 12-DEC-1996.  
 PF 06-JUN-1996; U09449.  
 PR 06-JUN-1995; US-467504.  
 PA (MONS.) MONSANTO CO.  
 PI John M;  
 DR WPI: 97-042726/04.  
 PT Plant fibre-specific, developmentally regulated FbLate promoter -  
 PT useful for producing transgenic plants, esp. cotton, with altered  
 PT fibre properties  
 PS Claim 8; Page 55-56; 79pp; English.  
 CC cDNA clones A8 or FbLate-1 (T43361) and All or FbLate-2 (T43362)  
 CC correspond to RNAs prevalent in late development of cotton  
 CC fibers. They were isolated from a 23-day cotton fibre cDNA  
 CC library by screening with 24-day fibre cDNA. A8 and All are  
 CC partial clones of the FbLate 2-82A gene. They can be used to  
 CC identify FbLate promoters (see also T43360) useful for fibre-  
 CC specific expression of foreign proteins in transgenic plants, esp.  
 CC cotton fibre.  
 SQ Sequence 974 BP; 388 A; 161 C; 222 G; 203 T;

Query Match 3.84; Score 210.6; DB 1; Length 974;  
 Best Local Similarity 68.14; Pred. No. 3.6e-20;  
 Matches 340; Conservative 1; Mismatches 145; Indels 13; Gaps 3;

Qy 4311 acgaagaagcagcagagagctgtaatacgaagaagcagaatacaaacgacgaatgcac 4370  
 Db 446 AAGAAAAACCCGATTTCCCAATGGGAAAGCCCTAAGGCGACGAGAAACATAAAGCCG 505  
 Qy 4371 aagagtagtcaaaacttgagagagctggaatgcanaag-----aggaanaaanaaaac 4424  
 Db 506 AATATCCGAATACTAGTGAAGCAAGAAAACTAGTAGGAGTAAAGAACATAAACATG 565  
 Qy 4425 cctgcagacagcatgaagtagtaccagcagatcaacagatcaaaagagcaaaagtagtac 4484  
 Db 566 AGTTCCCAAGCATGAAAAAGAGAGGAGAAACCTGAGGAAGCATAGTACCTGAG 625  
 Qy 4485 agaagaagaactctgcagccgcccgaagattcttcgtagcgtctgcagccgcccgaattc 4544  
 Db 626 TGGGTAAAAAGCCTGAATGGCCGAGTCCATGTTTCTCAGTCTGGCTCAG-----C 679  
 Qy 4545 gtgcagagctggaatcatgagcgtgtgcatgtgcatcatcatcgatgaatttcattg 4604  
 Db 680 ACTAGGCTTAAGCCATATGACATGTCATGTCATCATCATCATCATGATTTTCATG 739  
 Qy 4605 tatatgtat-tatatgtatataaaaaagagtggttggtgaatgtgtgtgattc 4663  
 Db 740 GATATGTTATATATATTGTTATATAAAAAAGAGTGGTGGGAATGTGTGTGTCATT 799  
 Qy 4664 ctccatgcactaatgtgtaattcttgcatacaataagaattcaaatggttatgttta 4723  
 Db 800 ATCCATGTCACATGCTGAATCTTTTGGATGACAGAGATCTGATGTTATATGTTTA 859  
 Qy 4724 tgttatgtatgttgttagtgaaktaatttaattgaattgtgtatgttaaatcaac 4783  
 Db 860 TGTTATATCTGTTGTCTAGTGAATAATTTGTAATGTTGTATGTATGTATATACATC 919  
 Qy 4784 ttggtgtgatttatgttat 4802  
 Db 920 TGGCTGTGATTATGTTTT 938

RESULT 12  
 T43362  
 ID T43362 standard; cDNA; 645 BP.  
 AC T43362;  
 DT 11-MAR-1997 (first entry)  
 DE Cotton FbLate 2-82A gene cDNA clone All (FbLate-2).

QY 4124 aaatacagctttttctttctattgataaacactgctctatgactgcttgcaccttc 4183  
Db 13 ACTAATAAAATCTTGCTTGCTTCTATTGTGTAACAACTGGCTATACACTTTGTCACTCTTC 72

QY 4184 tctcttttcccaactttctactatagtgctctactagtcacgctgtagccacactttgct 4243  
Db 73 TCTCCTTTTTCACACTTTTACTACTATCTGCTCACTAAATCGGTAGTCAACACGCTGCTC 132

QY 4244 gcagcgctgcagctgttatcgagacaacagcaactctcagagctccacacattgctt 4303  
Db 133 TCAAGCGCTGCACATTTATTCTCAGACACAACCACTCTCATCAGAGCTGCACAAATTTGCT 192

QY 4304 tcaaatatcgaaagcagcagagctctgaatacgaaagccgaatatacaaacgcgaag 4363  
Db 193 TCAAAATAGAGTACCAACAAAGCTCT-----GATCACTACCAACCAAAA 237

QY 4364 tctcagcagagtagtcaactcagctgaactcgaagcgaaggaagaaacacaa 4423  
Db 238 TATCAAGAAAGTATCCACAACACTGACAGCCCTTAATGCACAGGAGGAAACACA 297

QY 4424 ccttgcaacgcgtacagagctacccgcgtacacagatcaacagagcaaaagagatc 4483  
Db 298 CCGTCGAACACTCATGAGAGTACACAGATCACGCAAGTCAAGGAGCAACGAAGATC 357

Dn 121 TTTATTTGGGTTTGGGTTTGTGAGTTTTAGATAATTAATTAATTCGCAT 180  
Qy 1918 aattttctgtatttgaagaagtgctgaatttttttccaaattgaacgttcaaga 1977  
Db 181 AATTTTCTGTTATTGAAAGGATGTCGAATTTTTTCCAAATGAACGTTTANGA 240  
Qy 1978 attttctacgtgaattcgaataagtaagttgttttttagaagaattaaatagtt 2037  
Db 241 ATTTTCTACTCGAAATCAGAAATAGTGAATTTGTTTTAGAAAGTTAAATAGTT 300  
Qy 2038 agtatacgaatttttagttgtattgttgggaagaattagttatttttgaacaataat 2097  
Db 301 AGTATTCGATTTTAACTTTGATTTGGTGGAAAGTAATGTATGTTTTGACACATATAT 360  
Qy 2098 ttgacaataataaagttttctcaggaataaagcaaatctcttcttttttgaaaa 2156  
Db 361 TGCACATTAATTAAGTTTCTAGAAATAACGGAATATCTCTTTTTTTTTGAAAA 420  
Qy 2157 ttactaatgcaaaccaaacacggttttggggcgaaataactagcttaagtagtcag 2216  
Db 421 TTACTATCTCAGACCAACACAGTTTTGGGAAGCAATAATCTAGCTTTAAGTAGCAG 480  
Qy 2217 tgaactctcaaaactcgtctaaactctaggtcagtttgcgtgcctacagtagtaag 2276  
Db 481 TGTAATCTCAAAATCTGGTCATACTTCTAGGCTGAGTTTCTGTGCTACAGTAGTAG 540  
Qy 2277 tctatagaacttacctgcacaaacgaatgaagtcaggtcgaattctcaacttttct 2336  
Db 541 TCTATGAAATTTACTGCACAAAGCATGAGCTGAGGTGGAATCTACAACTTTCTCT 600  
Qy 2337 ttttctcaataacatattggttgatcgaattcgcgtatcaataattattacagttt 2396  
Db 601 TTTCTTCAATTAACATATGTTGATTCAAGTTCGGATCTAATAAATTTATTAGATTT 660  
Qy 2397 atcaattcttaactcttatctatctctataataaataagtcagttcaaatcagttt 2456  
Db 661 ATCAATTTCAATTAACATATGTTGATTCAAGTTCGGATCTAATAAATTTATTAGATTT 720  
Qy 2457 cgaagttcccaaaatttgaattttataaattttatctcaaaacgaataagttat 2516  
Db 721 GGAAGTTCCCAAAATTTGAATTTTAAATTTATCCCTAAAACGAATAGTAG 780  
Qy 2517 atcttcaaaatttaagtttcaattttcaatcgaatttcaatttcaatttcttataactct 2576  
Db 781 ATCTTCAAAATTAAGTTTCATTTTCAATCCGATTTCAATTCATCTTTTATAGCTCT 840  
Qy 2577 ctattatctataataacataaatttcaaaatttttgaattatttcaacttttagtctct 2636  
Db 841 CTAGATCTATATATACAAATTTCAAACTAATTTGAAATATATACATTTAGTCTCT 900  
Qy 2637 aagttcaaaactataaattttcattttagaatttaacttttcaatctatgaagtcga 2696  
Db 901 AAGTTCAAACTATAAATTTTCACTTTTGAATTAATCATTTTTCACATCTAGATCA 960  
Qy 2697 atttaaccaaatgcacaaatttctgatttagtagtaacagcttttagtcttcaaac 2756  
Db 961 ATTACCAATGACACAAATTTCTGATTAGTAGTAAGCAAGCTTTGAGCTTCAAAA 1020  
Qy 2757 ataaaatt----acaaaaaaacaaactaaactatttcaatattgaacacaaa 2812  
Db 1021 CATAAAAATACAAAAAACAACAACTTAAATCATTTATCAATTTGACACAAA 1080  
Qy 2813 gcttggcgaatgtagaagcttaaaatgctcttttcttcttttgggtgaacgg 2872  
Db 1081 GCTTGGCGAATGCTAAGAGCTTAAAGTGCCTTTTGTGTCTTTTGTGCAACGG 1140  
Qy 2873 tggagagagagggaatgaagatgacactatttttttatttatttgaactataata 2932  
Db 1141 TGGAGAGAGAGGGAATGAAGATGACCAATTTTTTATATGTTTTAACATATAATA 1200  
Qy 2933 ttaataatttaataataataatttcttggtaagtgaacgtgggagtagtaagta 2992  
Db 1201 TTATATTTAATCTATATTATCTTGTGTAATGTGACAGTGGGAGATCTAAAGTA 1260

Qy 2993 ttttaacattatctttttgcaagcagttggcgtctacccaagagtgatcaaaagttg 3052  
Db 1261 -TATACATTTATATCTTTTGGACAGCAGTTGGCGTCTACGAGAGTGATCAAAAGTTTC 1319  
Qy 3053 agcgcgtctcaatgagcacaattttggccatgatgtaaaagcaatttggtagtcaa 3112  
Db 1320 AGCTGCTCTCAATGGCCAAATTTTGGCCATATGGAATGAAGCAATTTGTTAGTTCAC 1379  
Qy 3113 ctgctcagcagataagttttaaattgaattaaattgaagtcgctcgtccacacacac 3169  
Db 1380 CTGCTCACGCAATAATGTTAAATGAATTAATTAAGTGCCCTGCACACACACAC 1439  
Qy 3170 aaaaaaaactaattgcttgggttgatttatataacggaatgaattattatttttt 3229  
Db 1440 AAAAAAAATCAATGTTGGTGGTGAATTTATATTAAGCAATGAATGTTATATTTA 1499  
Qy 3230 aaaaaaaattgtttttagcttctaatttttggagacttccatctataatttctgt 3289  
Db 1500 AATAAATTAATGTTATTTAGATTTCTTAATTTT--GAGCATCTCATTAATCTGT 1558  
Qy 3290 a-acataattataaataatgataataaagtgtaattaaattttaaatacaagcaaat 3348  
Db 1559 ATACATAATATAAAATATAGTATAATAAGTGAATTAAGTTTAAATACAGCATAA 1618  
Qy 3349 attaaatttgaatcaattatttttttcttattatttttaatttaatttagtttttt 3408  
Db 1619 ATTAAATTTGAATCAATTAATTTTAAATTTTCTATTATTATTAAATTTAGTCTATT 1678  
Qy 3409 tcaaaataaattttaaactcaaaaaaataatttttcttaattgtagcaactcag 3468  
Db 1679 TCAAAATAAATTTAAATCTAAATAAAATTAATTTTCTCTTAATTT----- 1725  
Qy 3469 ttatacttcaaaattataagttattatttactcgtgattttatttattagtatattaa 3528  
Db 1726 -----ATTAAATTTTATTTCACATATATTTTACTTATTAATATATAA 1773  
Qy 3529 tctgtattattatttgggtgagacagctcttccacaaatttttaactatgatt 3588  
Db 1774 TTAT----- 1777  
Qy 3589 ataaattttatccaactcgttatttacttataataataattttatcaattttatg 3648  
Db 1778 -----AATAATTTATCATATTTTATC 1799  
Qy 3649 gaattgagacagaagaacataagagacaaattctataaagaacatttgaaaaaa 3708  
Db 1800 GAAATTGAGACAGAAACATTAGAGACAAATTTCTAAGACAACTATTAG-TAAA 1858  
Qy 3709 aagtacttttagttatttttaagtaactcttaacaaacacaaaattcaactcaatga 3768  
Db 1859 AATGACTTTTAGTAAATTTTAGTACTCTTAACAAACACAAATTTCAATCAATGA 1918  
Qy 3769 actaaatagaataataacacagcaactcttctgtatttacttccacttccactt 3828  
Db 1919 ACCAAATAGATATAATACATACAGATATCTCTGTTGTTTCTACATCCCGTATC 1978  
Qy 3829 ttattatgaataaactcttattatctcgaactaaagtgttgcacaaattattatctaa 3888  
Db 1979 ATATTGAAAGTAATATTATATTACTGAGCAATATCTCTGTTGTTTCTACATCCCGTATC 2038  
Qy 3889 ataaaga--aaactaatttttataaactttttctatttttgaagattatttt 3946  
Db 2039 AAAAAAATGTTGAATTAATTTTATACATTTTTTCTATATTTCGAGATTTATATT 2098  
Qy 3947 tgtattatttagtaaaattttgacatagattgagcctctctacataactccacac 4006  
Db 2099 TGTATTTTACGTAAAAATTTTGACATAGATTGACACCTTCTACATACATCCACAC 2158  
Qy 4007 taagtcaagtttagatagaaatttgtaacaaacagctggggcaaatccacacac 4066  
Db 2159 TAGTCAAGTATGATGAGAAATGTGTACAAACAGCTGGGCGCAATCCACCAAC 2218

Db 2041 |||||TAGTATATACGATTTT|||TAGTTGTTGCTGGAAGTATGATGTTT|||TGACATA 2100  
Qy 2093 attatttgacataaataagttttctagggaataaaccggaataattctttctttttgt 2152  
Db 2101 ATATTATGCAATAATTAAGTTTCTAGGGAATAACGGAATATCTCTCTTTTGTG 2160  
Qy 2153 aaaattactactgacgaacaacaacgctttggggagcaataactctagcttttaagtag 2212  
Db 2161 AAAATTAATAATGCAAGAACAAACACGCTTTGGGAGCAAAATATCTAGCTTTAAGTAG 2220  
Qy 2213 toagtgtaactctcaaaattctggctataactcttagctgagtttgctgctcaagtag 2272  
Db 2221 TCAGTGTAACCTCAAAATCTGCTAATCTCTAGCTGAGTTGCTGCTCAGTAG 2280  
Qy 2273 taagtctatagaacttaactgacaaacgacatgactgaggtcgactcaacattt 2332  
Db 2281 TAGTCTATAGAACTTAACCTGACAAACGACATGACGTGAGGTGAAATACACATT 2340  
Qy 2333 tctcttttctctcaataaactatgcttgattgaatttcgactatataaattatttag 2392  
Db 2341 TCTTTTCTCTCAATTAACATATGTTGATTCAGATTCGATCATTAATTAATTATAG 2400  
Qy 2393 atttataactttcaatacttatactatctattataaataaagtcagtcacttag 2452  
Db 2401 ATTTATCAATTTCAATTAACCTATATCATCTTAATAAATAAGTCAGTCAATTCAG 2460  
Qy 2453 ttttgaagttcccaaaaatttgaattttataaattttccctaaaaacgaatat 2512  
Db 2461 TTTTGGAAGTCCCAAAATTTGAATTTTAAATTTTATTCCTGAACCAAGATAG 2520  
Qy 2513 ttatatctttcaatttaagtttcaattttcaactgacttaatttctcttttttaa 2572  
Db 2521 TATATCTTTCAATTTAAGTTTCATTTTCAATCGATTTCAATTCATCTTTATATA 2580  
Qy 2573 ctctctctctctataaattacataaatttcaaatatttgaattttacacttttag 2632  
Db 2581 CTTCTATATCATATATACATAAATTTCAATTAATTTTGAATTTTACATTTAGT 2640  
Qy 2633 ccttaagttcaaaactataaattttcctttgaataatctacttttccatcaatga 2692  
Db 2641 CCTAGTCTCAAACTATAAATTTCACTTTAGATAATTAATTTTCACTCAATGACA 2700  
Qy 2693 tcaaatttaacaaatgacaaacttctgattgtagtagtaacgcttttgagttct 2752  
Db 2701 TCAAAATTTACCAATATGACAAATTTCAATGATTTAGATCAAGCTTTGAGCTTCA 2760  
Qy 2753 aaacataaaattacaaaaaaacaaacttaaaactattatcaaatgacacaaa 2812  
Db 2761 AATCATTAATATACAAAAAAACAACTTAATATCATTTATCAATTTGACACAAA 2820  
Qy 2813 gcttgccgagatgtagaagcttaaaatgctctttgtttcttttttctgcaaacg 2872  
Db 2821 CTTGCGGATGCTAAGAGCTTAATAATGCTCTTTGTTCTTTTGTGCAACG 2880  
Qy 2873 tggagagagagaggaatgaagattgacataatttttattatggttttaacataata 2932  
Db 2881 TGGAGAGAGAGGGAATGAGATTGACATATTTTTTATATGTTTACATATATA 2940  
Qy 2933 ttaataatttaactcaataacttttgtagtagtagtagtagtagtagtagtagtag 2992  
Db 2941 TTAATAATTTATCATATATATCTTTGAGATGTCACATGCGGAGATAGTAAGTA 3000  
Qy 2993 ttttaacattatcttttgcaagcagttgctgctacacaaaggtatcaaatgtt 3052  
Db 3001 TTTTACATTAATCTTTTGCACAGATTTGCTGCTCACCAGAGTGATCAAGTTG 3060  
Qy 3053 agctgcttctcaagacaaattttgcccataatgataaaggaatttggtagtcaa 3112  
Db 3061 AGCTGCTCTCAATGACCAATTTTGCCATATGATAAGAGCAATTTGTTAGTCAA 3120  
Qy 3113 ctgctcagaataaagttaaatgaatttaaaatgaagtggtgctgctacacacaaa 3172

Db 3121 CTGCTCAGATAAGTTAAATGAATTAATAAGTGGCTGTCACACACAAA 3180  
Qy 3173 aaaaactaatgttgggttggaattttatataccggaatgaataattattttaaaa 3232  
Db 3181 AAAAAATAAGTGGTGGTGAATTTATATTACGAATGTATATTAATTTAAA 3240  
Qy 3233 taaaattatgttttagatcttaataattttggagcttccactataattctgtaac 3292  
Db 3241 TAAATATATATTATAGATCTTAATTTTGGAGCAATTCATATTAATTTGTAAC 3300  
Qy 3293 ataatataaaaatagataataaagtgtaatttaacttttaataacagataatata 3352  
Db 3301 ATAAATATAAATATAGTATATAAGGTAAATTAACCTTTAAATACAGCATATATA 3360  
Qy 3353 aattttgaactaattttttttttcttattatttttaattatttagtctatttttcaa 3412  
Db 3361 AATTATGAATCAATTAATTTTATTTCTATTATTTAATTAATTAGTCTATTTTCAA 3420  
Qy 3413 aataaatttaacttaaaaaaattttcttctaatgttgaaacactctgttat 3472  
Db 3421 AATAAATTTAAATCAATAAATAATTTTCTCTATGTGAACACATAGTTAT 3480  
Qy 3473 acttcaaaataataagattatatttactctgattattttattagatatattct 3532  
Db 3481 ACTTCAAAATTAAGATATATATTTTACCTGATGATTATTTATAGTATATTATCT 3540  
Qy 3533 gattataatttggtgggatacaatgctttcactaaatttttaactgattttata 3592  
Db 3541 GATTATATATGTTGGGATACATCGCTTCCATTAATATTTAACATGATTATATA 3600  
Qy 3593 attttattcaactgattatttacttataatacataatttctataattttatgaa 3652  
Db 3601 ATTTATTCACATCTATTTTACTTATATACATAATTTACATATATTATGAA 3660  
Qy 3653 tggcagaacgaacacttaagacaaacttctataaagaacacttagaaaaaatg 3712  
Db 3661 TGGAGCAGAGAACATTAAGAGACAAATCTATAACAGAACATTTAGAAAAAATG 3720  
Qy 3713 tcttttaggtatttttaagctacttcaacaaacaaaaattcaaatgaacta 3772  
Db 3721 TACTTTTAGGTATTTTAACTACTCTTAACAAACAAAAATTCATCAATGAACATA 3780  
Qy 3773 aataagataataaactcaggaactcttactgttaatttcaacttcccaataattat 3832  
Db 3781 AATAAGATATATAACATCGGAACATCTTACTTGTAACTTACATCCCAATTTTAT 3840  
Qy 3833 tatgaaaaataatttatactactggaactaaagtgttgcaaaatttatctaaata 3892  
Db 3841 TATGAAAAATATCTTATATCTGACATTAATGTTGTCACAAATATTATCTAAATA 3900  
Qy 3893 agaaaaacattatttttataaactttttctataattttgaagatttatatttgata 3952  
Db 3901 AGAAAAACCTTAATTTTATAACATTTTTCATTAATTTGAAAGATTATATTGTATA 3960  
Qy 3953 ttttagaaaaattttgacagattgagcaacttcttaacataatcccaaatgaact 4012  
Db 3961 TTTAGTAAAAATTTGACATAGATTGACACTCTTACATATATCCCAACCATAGCT 4020  
Qy 4013 agtatagtagtaggaattgtgacaaacagctggggcacaattcccaaacacatct 4072  
Db 4021 AGTATGTAGTAGGAATTTGGTCAACACAGCTGGGGCAATCCCAACACCATCTC 4080  
Qy 4073 tcttctctctctataaaggtgtgctacatagacacataccacacaaatcaact 4132  
Db 4081 TCACTCTCTCTATAAAGGCTTGCTACATAGACACATCCACACAAATACAGCT 4140  
Qy 4133 tctttttctctttttagtaaacctggctctagcattgtctccctttcttctttt 4192  
Db 4141 TCTTTCTCTTATTTGATTAACATGGCTATAGCACTGCTGCACTCTTCTCTTTT 4200  
Qy 4193 caacttttctctaaagtctctactgtagcagcttagcacaactgttctggcagcgt 4252  
Db 4201 CAACTTTACTCATAGTGTCTCACTAGTACCGGTAGCCACACTGTTTGGACGCGCT 4260

Qy 4561 tatgaagctgtgcatgtgcacatcatcgaatatttcgttatgatgaatata 4620  
 Db 4561 TATGACCGTGGTCATGTGCCATCATCAGCAATTTTCATGSTATATGTAATATA 4620

Qy 4621 gtaataaaaaagatgtgatgtgggaatgtgtgtgcatttcctcatgcactaatgt 4680  
 Db 4621 GTAAATAAAMAGATGTGATGGGAATGTGTGTGCATTCCTCATGCACATATGTG 4680

Qy 4681 gaattctttgatcacatgaataattcgaattgttatgttatgttatgttatgtt 4740  
 Db 4681 GAATCTCTTGCATACATAGAATTTCTAAATGTTATGATTTATGTATATGTATGTG 4740

Qy 4741 tagtgaaktaatttttaagtgttatcatatgaatcaactgcttggttatgtt 4800  
 Db 4741 TAGTGAATAATTTTAAAGTGTATCTAATGTTACATCACTTGGCTGATTTATGT 4800

Qy 4801 atgttatgtattttacttaataatgatgtgatgttatgaattcaactgtgatca 4860  
 Db 4801 ATGTATGTATTTTACTTTAATGATATGCAATGATTTGTAATTAACATGCTTGATCA 4860

Qy 4861 ttatactcttcatataattataaagtgaactgtttgttttaaaatttttaacagtt 4920  
 Db 4861 TTATACTCTTCTACTATTAAATATAAGGCACATGTTGTTTAAATCTTTTACAGTTA 4920

Qy 4921 agacagtataaataatgatcaaatataatacaagtttttagttcaagttagctatcta 4980  
 Db 4921 AGACATGTATAAATATATGACAATATAATACAAGTTTATGTTCAATGTAGCTATCTTA 4980

Qy 4981 gtatgttatgtgatcttctaattacattaaacaattccaattaaatttttaaaata 5040  
 Db 4981 GTATGTATTATGATGATCTTAATACATTAAACAAATTTCACTAAATTTTAAATA 5040

Qy 5041 atacaacaataatttattgataataatcaattgaacacaaataaataaataa 5100  
 Db 5041 ATACAATAATATTATTGTATATATAATACATTAATGCAACAAAATGAATAAATAA 5100

Qy 5101 taaatagcaataattgttataatattgtaataataatgtacacatttcttaactgaa 5160  
 Db 5101 TAAATAGCAAAATATTGTATAATATTGTAAATAATATGACCAATTCTTAACAGAA 5160

Qy 5161 atagggtctaaactataatcccaaaatttcagtttaaatattttataactgcatatt 5220  
 Db 5161 ATAGGCTCTAACTATATAATCCATAAAATTCAGTTTAAATTTTATTAACCTGCATATT 5220

Qy 5221 attagaactctttttaaataattaaattttataacacaaatttaattaaactatta 5280  
 Db 5221 ATTAGAATCTTTTAAATATATTAAATTTTAAATATACCAATTTAATTAAACTATTA 5280

Qy 5281 attatcttaactaaataactaaattttatttaacatttaataacttaacttaatt 5340  
 Db 5281 ATTATCTTAACATAAACTAAATTTTATTACATTAATTAATTAATCTTAATCTCT 5340

Qy 5341 atataatttaaaacttaattatcatattgatttaatttctgattatcttaattgt 5400  
 Db 5341 ATCTAATTAAACCTTAATATCCTAATTGATTTAAATCTTGAATCTTAATCTTAATT 5400

Qy 5401 aactctctccacacagctagatgtggaacacagatccggagatttaactggaatgaga 5460  
 Db 5401 AACCTCCCTCCACAGCTAGATGCTGACCCGAATCCGGAGATTACATGGCATGTGGA 5460

Qy 5461 tggcctgatgatgacaggtttttctagaggtaccacattgocctatgatgactgt 5518  
 Db 5461 TGCCCTAGTAGATCAGGGTTTCTAGAGGTACCAATGGCCCTATAGTAGTGCT 5518

OS Gossypium hirsutum cv. coker 130.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..65  
 FT /\*tag= a  
 FT /\*note= "fragment of pBluescriptII polylinker (as stated in the specification)"  
 FT misc\_feature 57..5494  
 FT /\*tag= b  
 FT /\*note= "genomic clone 4-4(6) from lambda phage clone of a cotton Coker 130 genomic library (as stated in the specification)"  
 FT misc\_RNA 65..4163  
 FT /\*tag= c  
 FT /\*note= "5' flanking region of the 4-4(6) gene (as stated in the specification)"  
 FT CDS 4163..4502  
 FT /\*tag= d  
 FT /\*note= "corresponds to part of the 4-4(6) ORF (as stated in the specification)"  
 FT CDS complement(4131..4502)  
 FT /\*tag= 1  
 FT /\*trans\_except= (pos:4170..4172, aa:Xaa)  
 FT /\*trans\_except= (pos:4182..4184, aa:Xaa)  
 FT /\*note= "Xaa = stop codon; No start or stop codons given, possibly conforms to exon structure. Encodes W21899"  
 FT misc\_feature 4502..4555  
 FT /\*tag= e  
 FT /\*note= "synthetic polylinker oligonucleotide containing unique target sites for EcoRI, SmaI, SalI, XbaI and BglII"  
 FT misc\_feature 4163..4555  
 FT /\*tag= f  
 FT /\*note= "stuffer fragment left in place to facilitate the monitoring of cloning manipulations (as stated in the specification)"  
 FT 3'UTR 4555..5494  
 FT /\*tag= g  
 FT /\*note= "corresponds to the 940 nucleotides downstream of the stop codon and constitutes the 3' flanking region of the 4-4(6) gene (as stated in the specification)"  
 FT misc\_feature 5494..5547  
 FT /\*tag= h  
 FT /\*note= "fragment of pBluescriptII polylinker (as stated in the specification)"  
 PN W09640924-A2.  
 PD 19-DEC-1996.  
 PF 07-JUN-1996; U09897.  
 PF 07-JUN-1995; U5-480178.  
 PR 01-JUL-1996; J2-005572.  
 PA (CAL) J. CALGENE INC.  
 PI McBride K, Pear JR, Perez-Grau L, Stalker DM;  
 DR WPI: 97-02523/05.  
 DR P-PSDB: W21899.  
 PT DNA construct contg. gene of interest controlled by cotton fibre transcriptional factor - used to produce altered phenotype cotton fibre cells expressing genes affecting pigmentation  
 PS Claim 22; Fig 2A-J; 95pp; English.  
 CC The present sequence is a 4-4 cotton fibre expression cassette (version 1) from promoter construct pCGN5606. The lambda genomic phase clone used to form this construct was designated 4-4(6). DNA constructs containing cotton fibre-specific transcriptional factor promoters are useful to produce cotton fibre cells with altered phenotype, especially altered colour. Genes involved in the production of melanin (e.g. tyrosinase gene and ORF438 encoded protein from Streptomyces antibioticus) and indigo (mono-oxygenase genes possibly in conjunction with a cytochrome P450 gene) are of interest. The promoters of the invention are reliable and permit expression of a protein selectively in cotton fibre to affect qualities such as fibre strength, length, colour and dyability as required. The construct and methods can also be used for the introduction of other advantageous genes into a cotton plant, e.g. a plant hormone. In particular, fibres from a plant producing coloured

RESULT 2  
 T73865  
 ID T73865 standard; DNA; 5547 BP.  
 AC T73865;  
 DT 26-JAN-1998 (first entry)  
 DE Cotton fibre promoter clone 4-4(6) construct, pCGN5606 (Version 1).  
 KW promoter; fibre-specific; transcriptional factor; promoter;  
 KW altered phenotype; colour; melanin; indigo; ss.



Qy 181 ctgtttacatctcttatatagctgaactacaacacttttagctaaaaaattggata 240  
Db 181 CTGATTTCACATCCCTTTATNAGGCTGAACACACAACTTACGTAAAAAATAGGATA 240  
Qy 241 acctaatgacaatatcaaatcagatatataaacactgatttttagctaacactttaacac 300  
Db 241 ACTAATGCAAAATCACAATCAGATATTAAACCATGATTTTACGTACNTTACACAC 300  
Qy 301 tttattgaaataaattgaaatttcaatctgctgatactgcccagaatttttaggcctaa 360  
Db 301 TTTATTGAAATCAATTGAAATTTCATCTGCTGATATGCCAAGAATTTAGGCCCTAA 360  
Qy 361 cagatttgggtggaactttaacatgcatgcatgttaactgtttgaaacaaagtttttt 420  
Db 361 CCGATTGGTGGTGAACCTTAACATGTCATGCATTTGTACCTGTTTGAACAAAGTTT 420  
Qy 421 goattatttactatatagaactgttttagtttagttgagtaaacacatgactgttaagct 480  
Db 421 GCATATTATTACTATATGAACGTGTTGATTAGTGTGAGTTACACACTGAGCTGTGAAGT 480  
Qy 481 cactcaaatttttctaaatttcaagtgatcgaacaaacttagcagggcggtgcagag 540  
Db 481 CACTCAAAATTTTCTAATTTCTAAGGTATCAGCAAACTTAGAGACGGCGGGCTAGAG 540  
Qy 541 agctcggattgattttctatgtaataaataagcatttattgtttttaaactattatgga 600  
Db 541 AGCTCGGATTGATTTCTAGTAAATAATAGACAGATTATGTTTAACTATTATGGA 600  
Qy 601 ctitttggactatgaactgtttgggaactttattttgtttttttatttggcttttttga 660  
Db 601 CTTTGGGACTATGTAACGTGTTGGGACTTTATTTTGTTTATTATGCTTTTGGGA 660  
Qy 661 tttagtaattattttttaaactgcaaaattatagtttttaaacaactaagtcacagt 720  
Db 661 TTTAGTAATTATTATTTTAACTGCAAAATATTATGTTTAAACAACATAGTCACAGT 720  
Qy 721 ttcaaaattccataactagaaatttttgcgtcaaaataaagtaactcatttaagtgctt 780  
Db 721 TTCAAAATTCCTAACTAGTAATTTTGCCTGCAAAATTAAGTAATCATTTAAGTGTT 780  
Qy 781 ttctgtaataaataaataaataattttaaagcagattttctcaaaattggaattgat 840  
Db 781 TTCTGTAATAAATAAATAAATAATTTTACGAGTATTTCCTAAAATTGGAAATGTAT 840  
Qy 841 ttacaaaatttagtatgtcacaacacagcttttatatgtcacaggcgatattcgtatggc 900  
Db 841 TTACAAAATTAGTATGTCAAAACACATGTTTATATGTTACAGCGGATATCCTCTAGGC 900  
Qy 901 aaataacatcagcggggttttgagtggttacaggcgagtggtgctatttttgatgaagt 960  
Db 901 AAATACATCAGCGGGGTTTGAGTGTTCACAGCGCAGTGGGCTCATTTTGAGTAGT 960  
Qy 961 atagttagggccgagtttttagattgcatcttaacagtcgaagattttgtaacttcgatg 1020  
Db 961 ATAGTTAGGCGCCAGTTTAGATTGCAATATCAGGCTCAAGATTGTTGAACCTCGATG 1020  
Qy 1021 aatgatatgtatgatgttcgatttaacgaataattgttttttcttttctgtgtgttttat 1080  
Db 1021 AATGATATGTATGATGTCCGATTAAGCAATATGTTTTTCTTTCTGTGTGTGTTAT 1080  
Qy 1081 ctgctgttagaattatatagtatttttctcaactcttataggaatgacattgtggc 1140  
Db 1081 CTCGTGTGATAGTATATGATGTTTATCCAACTCTTAGGCATGTGACATGTGGC 1140  
Qy 1141 tattctaaattaaattgattgttatattgaactgtgcatgctgttctcaaaagcatg 1200  
Db 1141 TATCTAATTAATGATTGTATTAATGAANCTGATGATGCTGTCTCAAAAGCATG 1200  
Qy 1201 gaactcatgctactgctttctgttaaagatcagtggaatttttaactgcttactat 1260  
Db 1201 GAATCTCATGCTACTGCTTTCTGTAAAGATGCAATGCAAGTTTACATGCTTACTAT 1260  
Qy 1261 ttgtattgtctcttgcatgctatgctacattacataggggtgggaatgatatggaaga 1320  
Db 1261 TTTGATTGTCCTGCTGCATGCTATGTACATATCATGGGGTGGGATGATATGGTAAGA 1320  
Qy 1321 ggaagtgttgacagtttaattgattgacatctgtggtgttaacacacattttgtatg 1380  
Db 1321 GGAAGTGTGACAGTTTAAATGATTGTCATATCTGGTGTTTAACACCAATTTGTTATG 1380  
Qy 1381 gcatctgtgctggttatgctgagctgagccacatctgtgtgtggaatttatctgt 1440  
Db 1381 GCATCTGACCTGGTATATGCTGCTGACCCGCAATCTGTCTGGAAATTTATCTGT 1440  
Qy 1441 gactctgtgctggttatgctacaaattttgtggtgtgtttgtgtaggaagctgtggg 1500  
Db 1441 GACTCTGCTGCAATGCTACAAATTATGTTGTGTGTTTGGATGACAGTCTGGG 1500  
Qy 1501 gaactctatttgggtgtgtgctgaggtgggtaggaaattttgaaaaaaatttgcattgt 1560  
Db 1501 GAACCTATTATGCTGTGTGGAGTTTGGTAGGAATTTTGAAAAAATTTGCAATGT 1560  
Qy 1561 gttttctgaaaaatttgatttaacataatcattgatttattttgttaattggaac 1620  
Db 1561 GTTTTCTGAAAAATATGCAATACAAATATGATATCTTCAATTTTGGCAATGAC 1620  
Qy 1621 gttataaaattctctatgatatctgactgtttattactatatgttttatgctga 1680  
Db 1621 GTTATAAAATCTCTATGATATCTGATCTGTTTATTAATATATGTGTTATGCTGA 1680  
Qy 1681 gttagctcaaacattgagattctatgctaccccaattatttaacttcaggcaactgt 1740  
Db 1681 GTTAAGTCAACATGAGATTATAGCTACACCAATTTTAAATCAATTTTCAAGCAATG 1740  
Qy 1741 cagacttaggattggaatgctgagagagctgaggtgttttctacacatattttat 1800  
Db 1741 CAGACTTAGGATTGAGTGGCTGAGAGAGCTGAGTGTGTTCTCAGATATTTTAT 1800  
Qy 1801 taataataatttaataaattatgaacttttggactgtcactaattttcgaattt 1860  
Db 1801 TAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860  
Qy 1861 tttattgttttgggttttggatttttttagataatttttaataattctgcataat 1920  
Db 1861 TATTGCTGTTGGGTTTTGTTGAATTTTATGATAATTTTAAATATCTGCAATAT 1920  
Qy 1921 ttttctgtatttgaagaagtgctgaatttttttcaaaattgaaagctttaagaatt 1980  
Db 1921 TTTCTGTATTGTAAGAGATGTGCAATTTTCTTCAAAATGAACGTTTAAAGATT 1980  
Qy 1981 ttactactgcaaatccgaataagtgaaatttttttagaagaatttaaatagttagt 2040  
Db 1981 TTACTACTGCAAAATCAGAAATGCAATTTGTTTGAAGAATTAATTAAGTTAGT 2040  
Qy 2041 atacgaatttttagttttagtttgggtgaaggaattgtttttgacataattttgt 2100  
Db 2041 ATTAGCAATTTTAGTTGATTGTTGGTGAAGTAATGATGTTTGAACATTAATTTG 2100  
Qy 2101 acaataattagtttctgaggaataaagcaaatattctcttttttgaataattac 2160  
Db 2101 ACAATAATTAGTTTCTAGGAATAACGGAATATCTTCTTTTGTAAATATAC 2160  
Qy 2161 taatgacagacaacaacagctttgggggcaataattctagcttaagtatcagtgta 2220  
Db 2161 TAATGACAGACAACAACAGCTTTGGGGACCAATATCTAGCTTTAAGTATGCAATGTA 2220  
Qy 2221 actctcaaaattgtgctcaactctagctgagtttggctgctgacagtagtaagctga 2280  
Db 2221 ACTCTCAAAATCTGTCTAACTCTAGGCTGATGCTGCTGCTCACTAGTAGTACTA 2280  
Qy 2281 tgaacacttaactgcaacaacagcagtcagctgagctgacacattttcttttt 2340  
Db 2281 TAGAATCTACTGCAAAACGACATGACCTGACAGGCTGAACTTCAACTTTCTTTT 2340  
Qy 2341 ctcaaatcaaatgttgatgataagtttcagatctataaattattacagattatca 2400  
Db 2341 CTTCAATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400